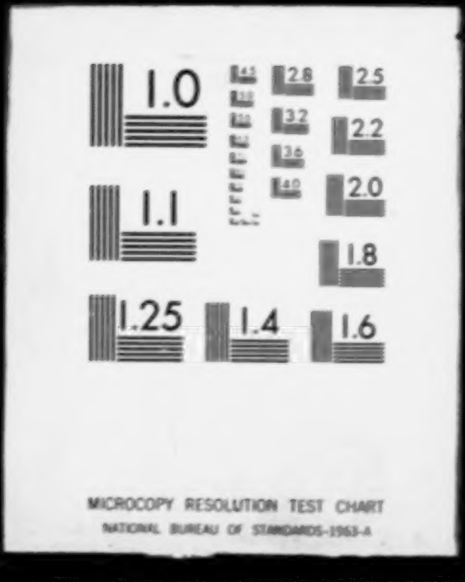
1 OF 3

ED 076 075


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

DOCUMENT RESUME

ED 076 075 FL 004 024

AUTHOR	Suppes, P.; And Others
TITLE	The French Syntax and Semantics of Philippe, Part I: Noun Phrases. Psychology & Education Series.
INSTITUTION	Stanford Univ., Calif. Inst. for Mathematical Studies in Social Science.
SPONS AGENCY	National Science Foundation, Washington, D.C.; Office of Naval Research, Arlington, Va.
REPORT NO	TR-195
PUB DATE	3 Nov 72
NOTE	196p.
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	Ambiguity; *Child Language; Form Classes (Languages); *French; *Generative Grammar; Grammar; *Language Development; *Language Research; Linguistic Performance; Linguistic Theory; Mathematical Linguistics; Nominals; Probability; Psycholinguistics; Research Methodology; Semantics; Speech; Statistical Analysis; Syntax; Tables (Data)

ABSTRACT

This is the first in a series of reports concerned with the analysis of a young child's spoken French. The research methodology is described and the corpus used for analysis is presented; the bulk of the analysis concerns the noun phrases used by the child. The main thrust of the report deals with the type of grammar applied to the corpus for analysis; the authors advocate the use of a probabilistic generative grammar which introduces probabilistic parameters which govern the use of a given generative rule and which permit a goodness-of-fit criterion which discriminates between the suitability of two different grammars for the same corpus. Two different grammars are applied to the corpus for analysis and the authors indicate a preference for the second grammar which, although it accounts for a slightly lower percentage of the total utterances in the corpus, provides a better fit in terms of frequency of uses of the generative rules matching the actual frequencies of grammatical types. (Author/VM)

FILMED FROM BEST AVAILABLE COPY

ED 076075

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

THE FRENCH SYNTAX AND SEMANTICS OF PHILIPPE,

PART I: NOUN PHRASES

by

P. Suppes, R. Smith and M. Léveillé

TECHNICAL REPORT NO. 195

November 3, 1972

PSYCHOLOGY AND EDUCATION SERIES

Reproduction in Whole or in Part Is Permitted for
Any Purpose of the United States Government

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

FL004024

ACKNOWLEDGMENTS

Support for this research was provided by the National Science Foundation under grant NSFGJ-443X and the Office of Naval Research under contract N0014-67-0112-0049. This research was performed in collaboration with the Laboratoire de Psychologie Expérimentale et Comparée associé au Centre National de la Recherche Scientifique, Paris, France.

TABLE OF CONTENTS

Section | Page

I. INTRODUCTION 1

II. COLLECTION OF THE CORPUS 6
  Choice of the child 6
  Conditions and frequency of recording sessions 7
  Transcription and editing 10

III. DICTIONARY I 14
  Articles 24
  Nouns 24
  Pronouns 25
  Adjectives 26
  Verbs 27

IV. NOUN-PHRASE GRAMMAR I 34
  Nonterminal vocabulary of Grammar I 37
  The production rules 40
  Nonprobabilistic fit of Grammar I 45
  Probabilistic fit of Grammar I 50
  Probabilistic lexical disambiguation 65
  Grammar I and the sections of the noun-phrase corpus 67

V. DICTIONARY II 74

VI. NOUN-PHRASE GRAMMAR II 78
  Nonterminal vocabulary of Grammar II 78
  The production rules 79
  Nonprobabilistic fit of Grammar II 85
  Probabilistic fit of Grammar II 87
  Probabilistic lexical disambiguation, Grammar II 101
  Grammar II and the sections of the noun-phrase corpus 102

TABLE OF CONTENTS (continued)

APPENDICES

APPENDIX A Analysis of Utterance Length 112

APPENDIX B Analysis of Noun-phrase Length 117

APPENDIX C Vocabulary of Philippe 122

APPENDIX D Dictionary I 153

References 191

I. INTRODUCTION

This is the first of a series of reports concerned with the analysis of a young child's spoken French. In the body of the paper, and also in the appendices, we give basic data on the corpus and how it was collected in Paris.

The main thrust of the report, however, is a detailed analysis of the corpus in the spirit of the work on probabilistic grammars and model-theoretic semantics begun several years ago in the Institute for Mathematical Studies in the Social Sciences at Stanford: Gammon (1970); Suppes (1970, 1971, 1972); and Smith (1972).

We begin only the first stage of analysis, namely, the analysis of the grammar of noun phrases in Philippe's speech taken from the first three hours, the middle three hours and the last three hours of the corpus, which ranges from the time that Philippe was 25 months old to 39 months old. In this and subsequent reports, we shall analyze the developmental aspects of Philippe's speech and shall investigate the grammar of the complete utterances, as well as the set-theoretical semantics of those utterances. However, to anticipate results reported in detail below, we did not find striking developmental trends in the grammatical structure of Philippe's noun phrases.

The kind of probabilistic generative grammar applied to the corpus represents the application of a certain set of concepts that we think are important in the detailed analysis of natural language. Some of our ideas are at variance with those commonly held by many linguists, especially the role we assign to probabilistic considerations. Consequently, even though these ideas have been set forth earlier in the publications mentioned above, it seems desirable to repeat some of our general arguments in support of the viewpoint we have adopted.

Many linguists deeply interested in generative grammars have very negative views about the kinds of frequency counts of vocabulary and utterance length that characterize statistical linguistics. We want to urge that there is no fundamental opposition between a generative viewpoint of grammar and a probabilistic analysis of grammatical types. What we propose is, we believe, a useful marriage of the two viewpoints.

By introducing probabilistic parameters that govern the use of a given generative rule, we are able at once by standard statistical methods to introduce a goodness-of-fit criterion to discriminate between two different grammars for the same corpus. It is important that this criterion is different from the criterion of which grammar accounts for a larger percentage of the utterances in the corpus. In the present paper, we consider two grammars. The first grammar accounts for a slightly higher percentage of the total utterances in the corpus, but from the standpoint of a probabilistic fit in terms of frequency of uses of the generative rules matching the actual frequencies of grammatical types, the second grammar is much better. It is also our feeling that the conceptual perusal of the two grammars will indicate why the second grammar is intellectually more satisfactory than the first. The requirement that the parameters attached to the generation rules match the actual frequencies with which these rules are used in derivations is to impose a criterion that requires the grammar to fit the corpus in a detailed way.

Those who have been concerned only about a competence grammar will not be deeply interested in our probabilistic results. We shall not examine in this paper the competence-performance issue, but simply say that we believe it is reasonable to think in terms of a performance grammar in dealing with a corpus of spoken speech and a probabilistic criterion is natural when dealing with a performance grammar.

We also investigate the possibility of introducing additional mechanisms on the generation of utterances, and we mirror these mechanisms in restrictions on the probabilistic parameters. In later sections we consider the mechanism that has the production rules for a given nonterminal symbol, for example the rewrite rules for determiners, be located in a push-down store. For this mechanism, a truncated geometric distribution governs the selection of which rule to use for a given generation. We recognize that this assumption is too simple, but it is interesting to see what decay in the goodness-of-fit results from applying this strong additional structural assumption.

Generally speaking, it is part of our program to consider strong additional assumptions and to match these assumptions with appropriate probabilistic parameters in order to obtain increasingly detailed and useful models of speech production.

Finally, we want to remark that in no sense do we consider the probabilistic account of the use of production rules an ultimate account. These probabilities themselves are overruled in particular cases by the semantics of a particular utterance. They represent the results of averaging over a number of utterances.

On the other hand we wish to emphasize that it is far from our belief that the introduction of semantics moves us from probabilistic to deterministic models. Probabilistic parameters are also needed at the level of semantics, for we are a very great distance from understanding the deterministic production of any nonstereotyped speech. In this report we consider the use of the parameters in dealing with ambiguity — i.e., the situation where there are two distinct analyses of a given utterance. We believe that the detailed examples worked out in later sections of the paper show how the probabilistic apparatus can be used in a constructive way to help in disambiguation. In subsequent reports dealing with the Philippe corpus we consider appropriate probabilistic parameters and their application to semantic disambiguation.

II. COLLECTING THE CORPUS

Choice of the child. The subject for our study was chosen for two reasons. First, we wanted to ensure that the child chosen would be in close association only with native French speakers. Second, we needed to find parents who would be willing to submit to the rather demanding rule that all they said, as well as what their child said, would be recorded for one hour a week over an indefinite period of time. Thus Philippe, the only child of a couple who were in their thirties and members of a university community, was chosen as our subject.

When we visited him for the first time, Philippe was 25 months and 19 days old. He was a sociable little boy who was not shy, even with strangers. During the period of data collection he often went to the Faculty of Sciences with his father who taught there. He also visited his mother in a laboratory of psychology where she worked, and he occasionally participated in experiments in the laboratory. Usually he attended nursery school; when he did not stay there the whole day a lady in her forties stayed with him at his house. Both his mother and father talked a lot with him and provided him with a verbally and intellectually stimulating environment.

Conditions and frequency of recording sessions. During the first period, April 22 through June 24, 1971, the observer (M. Le'veille') visited Philippe in his house one hour a week. A group of 10 sessions was completed by June 24, 1971, before summer vacation began. After an interruption of nearly three months (83 days), when Philippe went to the country, the sessions continued at the same frequency through December 18, 1971. There was a lapse of 14 days between September 30, 1971 and October 14, 1971 due to a strike on the Metro, which paralyzed Paris. At that time 21 hours had been recorded. Then the visits became less frequent -- one every fortnight. Again, between December 18 and January 6, 1972, and March 23 and May 6 a total of 63 days elapsed because Philippe was on vacation. Session 33 was the last one, because Philippe was leaving Paris for his summer vacation. The complete schedule of recording sessions is given in Table 1.

TABLE 1

Schedule of Recording Sessions

No.	Date	Age	
1	April 22, 1971	25 months	19 days
2	April 29, 1971	25	26
3	May 6, 1971	26	3
4	May 13, 1971	26	10
5	May 20, 1971	26	17
6	May 29, 1971	26	26

7	June 3, 1971	27	0
8	June 10, 1971	27	7
9	June 17, 1971	27	14
10	June 24, 1971	27	21
11	September 16, 1971	30	13
12	September 23, 1971	30	20
13	September 30, 1971	30	27
14	October 14, 1971	31	11
15	October 21, 1971	31	18
16	October 28, 1971	31	25
17	November 4, 1971	32	1
18	November 11, 1971	32	8
19	November 18, 1971	32	15
20	November 25, 1971	32	22
21	December 2, 1971	32	29
22	December 18, 1971	33	15
23	January 6, 1972	34	3
24	January 20, 1972	34	17
25	February 3, 1972	35	0
26	February 10, 1972*	35	7
27	February 24, 1972	35	21
28	March 9, 1972	36	6
29	March 23, 1972	36	20
30	May 6, 1972 (6)	38	3
31	May 18, 1972	38	15
32	June 1, 1972	38	29
33	June 15, 1972	39	12

*According to the regular schedule, Session 26 should have taken place on February 17, but because Philippe was scheduled for vacation on this date, the session was recorded earlier.

For practical reasons, recording sessions always took place in the morning, generally not long after Philippe had awakened. Each session, with a few exceptions, lasted one hour. Although the recording periods were relaxed and informal, Philippe was asked not to leave the room in which the tape recorder (a UHER Recorder L 4000) was installed for any significant time. Usually the tape recorder was set up in the living room, which was at the center of the apartment. Only the microphone was moved to the kitchen during breakfast. If Philippe wanted to play in his bedroom, the tape recorder was taken there, and Philippe was asked not to go into the other rooms too often or too long.

There were few variations in the sessions, because they were held at about the same hour and under the same general circumstances. After the observer arrived, breakfast was served in the kitchen. Following breakfast, Philippe would play or look at books with the father or the mother present, or possibly both, especially in the early sessions. The observer took extensive notes on the behavior of the participants.

During the first session the mother, who had explained in advance the reason why the observer came, showed Philippe how the tape recorder worked. Philippe became accustomed rapidly to the routine and accepted it without difficulty. There was no selected topic of conversation in any of the sessions. The parents as well as the observer encouraged Philippe to speak, and consequently, asked him many questions about his school activities and what he had seen or done during the week. At times Philippe was not interested in the conversation between his parents (mostly during breakfast), but he intervened in the discussion whenever he wished.

Transcription and editing. The transcription of the recordings was made by the observer, always within a week following the session. The transcription includes utterances of both the child and the adults. In addition, comments on the perceptual and social situation based on the notes taken during the session have been included in the transcription. Additional comments by the observer that facilitate the understanding of what Philippe said have also been inserted.

The transcription was made without use of a phonetic system. The rules of written French were applied, except when a spelling different from the standard one served the purpose of revealing a clear deviation from the usual rules of spoken French. For example, when Philippe pointed to a horse drawn in a book and said *un chevau*, the *x*, which is the correct ending of *chevaux*, was omitted in the transcription in order to differentiate Philippe's special usage and to lead to easy identification of this usage in the transcribed text. As a second example, Philippe said *des monsieurs* instead of *des messieurs*. A tabulation of these nonstandard words is given in the next section.

In contrast, the words incorrectly pronounced by Philippe but clearly identified by the observer have been transcribed with the standard spelling. For instance, Philippe often said *de'hors* instead of *dehors*. He also had difficulty in pronouncing correctly *cacher* and *casser*, but what he said was sufficiently clear to make his meaning understood.

The symbol &&& has been used to indicate unintelligibility of what was said by the child or by the adults present. This symbol may stand for a single word or for an entire utterance. Pauses in the conversation have been indicated by a new line, but in this as in other cases, the criterion for making the judgment was always somewhat subjective.

Because the transcription was made on a standard Model-33 teletype with ASCII characters, it was necessary to introduce certain character conventions for the transcription of French. The acute, grave, circumflex and diphthong accents were all replaced by an apostrophe. The apostrophe was replaced by the slash (/) and the cedilla by an asterisk. In the body of the present paper, however, a notation closer to the standard French accents is employed, but the above conventions were used in several of the tables taken directly from the computer printout.

The coding of accents created certain minor unanticipated problems. For example, since the apostrophe was replaced by a slash, some words which were written with an apostrophe, for example, *aujourd'hui* and *quelqu'un*, appear in the vocabulary as two words. Thus, _hui_ has been coded as an unclassifiable miscellaneous word in the vocabulary. In the case of _quelqu'un_, _quelqu_ has been coded with its proper code and _un_ as an article. Because the frequency of these examples is low, we did not redo the final analysis to take account of them.

In the final perusal of the dictionary, we noticed that three words were misspelled: _charriot_, _hibous_, and _ele'phant_. The word _c*_ should be _ç_. Moreover, the word _Oh_, which appears at the end of the vocabulary, is due to confusion between the numeral and the letter. Extensive computer reruns would have been required to correct these minor points, but these were not made, because no significant error is introduced by their retention.

As an additional remark, some words used in the corpus by the mother and father are provincial expressions and not standard French. They have been coded as standard words and will not appear as unusual words in subsequent analyses of the adult speech. This does not generate a problem for the tables of the present report because the vocabulary in this study is restricted entirely to Philippe's speech and the expressions of the adults are not pertinent.

For further information, the distribution of the length of utterances is discussed in Appendix A, and the distribution of the length of the noun-phrases is discussed in Appendix B. The vocabulary of Philippe's utterances in the 33 sessions is listed in Appendix C.

III. DICTIONARY I

The first step in writing a grammar and semantics of Philippe's utterances was to establish the grammatical categories for classification of terminal words. Basically, we followed classical morphological analysis and used the following eight classes of words: articles, nouns, pronouns, adjectives, verbs, adverbs, prepositions and conjunctions. Subclassifications within these broad classes are described below. Two additional classes were also used: onomatopoeias and interjections (KO), and uncodable expressions (MS), that is, words that were either not understandable or sounds that could not be identified as clearly representing a word. Some words that are acceptable in French were coded as MS since, in the context of their utterance, they were not words, but rather only sounds that Philippe repeated without understanding. The indistinguishable utterances classified as MS were not included in the analysis of the noun phrases. Examples include cor and tonne. In Tables 2 and 3 we show the mispronounced and fragmented words, and occurrences of nonstandard usages in Philippe's speech.

Appendix C gives a complete reference to the vocabulary of Philippe, and includes both rank and alphabetical orderings of the vocabulary. Appendix D gives the complete listing of the coding in Dictionary I.

TABLE 2

Mispronounced and Fragmented Words

Corpus	Interpretation	Code
baguitte	baguette	NC2.0
banguette	"	NC2.0
beguitte	"	NC2.0
be'toleu	be'tonneuse*	NC2.0
be'toleuse	"	NC2.0
blanquette	banquette	NC2.0
bouette	brouette	NC2.0
bousse	boule	NC2.0
cace	cache	VT1
chigne	cygne	NC1.0
codile	crocodile	NC1.0
drabon	dragon	NC1.0
e'dicaments	me'dicament	NC1.0
e'glique	e'glise	NC2.0
estatues	statues	NC2.0
fabouilles	farfouilles	VI1
festation	manifestation	NC2.0
kola	koala	NC1.0
magne'ophone	magne'tophone	NC1.0
mama	maman	NP2
mamaman	"	NP2
masou	mai's	NC1.0
migre	e'migre	VI1
modeur	moteur	NC1.0
plas	pas	DV6
positoire	suppositoire	NC1.0
retourer	retourner	VN0
samedie'	samedi	NC1.0
toffe	e'toffe	NC2.0
verture	couverture	NC2.0
xagone	hexagone	NC1.0

* be'tonneuse does not appear in "Le Petit Robert," but the parents used that word instead of be'tonnie're.

TABLE 3

Occurrences of Nonstandard Usage

Words	Code	Session	Typical context
alles	VI15,VA15	17	je suis pas content que tu t/en alles
aye	VT15,VA15	30	il faut...un soleil chez moi pour qui aye plus de vent
ayent	VT15,VA15	27	pour qu/ils ayent chaud
batter*	VTO		(it is said for battre)
chevals	NC1.2	10	...et les deux chevals
chevau	NC1.1	6	une chevau elle court
conduiteur	NC1.0	17	il a oublie' le feu rouge le conduiteur
e'clant*	NC1.0,VT1	2	(it is said for aimant)
e'crise	VT5	30	il faut que je t/e'crise quelque chose
loine	DV2,LCO	26	dans une maison tre's loine
monsieurs	NC1.2	12	c/est pas bien de tuer les monsieurs
morde	AQ2.0.2	33	j/ai jete' l/allumette qui e'tait morde
mordes	AQ2.0.2	28	les feuilles mordes
nallumettes	NC2.0	14	j/ai fait une grande maison avec des nallumettes
nanimals	NC1.2	20	des nanimals
narbre	NC1.1	19	il est noir ce narbre
nautres	AQ0.0.1,PI8,AI	3	des nautres rails
navion	NC1.0	9	voila' le petit navion
nen	PE19,DV2,E	27	c/est les enfants qui nen mangent
nenfant	NC0.0	21	c/est pas des petites te'tes de nenfant
noiseau*	NC1.0	3	(it is said for oiseau)
nours	NC1.0	1	tombe' le nours
oeils	NC1.2	17	c/est c*a les oeils?
prendait	VN6	26	elle prendait toutes les sous
prende	VN5	26	il e'tait pas content qu/on lui prende toutes ses sous
rate	VT1	28	on rate les feuilles mordes
se'que'	AQ1.0.2	19	tu veux un petit raisin se'que'?
se'que's	AQ1.0.2	19	tu veux des raisins se'que's?
soye	VT15,VA15	31	pour que la maison soye bien grande
soyent	VT15,VA15	16	...pour elles soyent plus belles
talle'	VIO,VAO	31	...je suis talle' a' l/e'cole
tuser*	VTO	12	(it is said for user)
zailes	NC2.0	13	il a pas de zailes
zarrivent	VI1	29	les voila' qui zarrivent...
ze'coutent	VT1	24	...et qui ze'coutent pas
zencore	DVO	36	y en a zencore?
zoiseau	NC1.0	3	un zoiseau sur le tonneau

zont	VT5,VA5	26	...c/est les chats qui zont
zoutils	NC1.0	12	c/est un carre' de zoutils

* Words associated with either an action, an object. Interpretation is based on the notes of the observer.

A standard problem in the construction of dictionaries for psycholinguistic corpora is that a given word may function in different ways depending upon its context. Thus, *avoir* may function as a transitive verb, as in *j'ai de l'argent*, or as an auxiliary, as in *j'ai fait mon travail*. Words of this kind have been assigned to two or more categories, and are signaled by commas in the coding. Transitive verbs are symbolized VT, and auxiliary verbs VA, so forms of *avoir* are classified VT,VA. When this occurs, we shall say that a word has been *multiply classified*.

We have not included either all or only those multiple classifications that are in all likelihood present in Philippe's speech. We have tended to be somewhat conservative. When it is highly unlikely that Philippe would have used a certain word in more than one context, we did not multiply classify that word. An example, the word *mets* has been coded as a verb but not a noun, since it is our belief that the latter usage ("something related with food") was unlikely to occur.

The subclassifications of the main classes were modified from classical French grammar as represented, for example, in Gre'visse (1969), although we did deviate from Gre'visse in a number of respects. In the first place, his and other classical grammars involve written language rather than spoken language. Especially in the case of verbs, the inflectional patterns have been extensively simplified to correspond to spoken rather than written inflections. After our analysis had been completed, we became aware of the work by Martinet (1958) and found that our analysis agrees to a large extent with his. Also, we have drawn upon the work of Dubois (1965) and Dubois-Charlier (1970), in particular Dubois' encoding of the definite and indefinite articles and the partitive constructions.

Because of the differences between written and spoken French, and because Gre'visse does not provide a completely systematic classification, it was necessary to introduce a number of distinctions not found in his analysis. Thus, a systematic coding was initiated which distinguishes whether the gender or number of an adjective or noun is undetermined, or whether the gender or number (singular or plural) is masculine or feminine. This is not to imply that Gre'visse ignored the fact that adjectives may have the same form regardless of whether they are masculine or feminine, but only that we felt it necessary to introduce clear coding conventions to describe these cases.

The coding introduced for each of the classes described above is exhibited in Tables 4 and 5. We turn now to detailed remarks about the nature of this classification and the reasons for it. To a large extent we have retained the standard French grammatical terminology in order to facilitate comparison with various published grammars of French.

TABLE 4

Coding of Grammatical Categories, Dictionary I

ARTICLE

Indefini

IN 1	un
IN 2	une

Defini

DN 1	le, l'
DN 2	la, l'
DN 3	les

PRONOM

Personnel

PE 1	je, j'	PE 11	lui
PE 2	me, m'	PE 12	eux
PE 3	moi	PE 13	se, s'
PE 4	nous	PE 14	soi
PE 5	tu, t'	PE 15	le, l'
PE 6	te, t'	PE 16	la, l'
PE 7	toi	PE 17	les
PE 8	vous	PE 18	leur
PE 9	il, ils	PE 19	en
PE 10	elle, elles	PE 20	y

Demonstratif

PD 1	celui
PD 2	celle, celles
PD 3	ceux
PD 4	celui-ci
PD 5	celui-la'
PD 6	celle-ci, celles-ci
PD 7	celle-la', celles-la'
PD 8	ceux-ci
PD 9	ceux-la
PD 10	ceci
PD 11	cela
PD 12	ca
PD 13	c'
PD 14	ce

Interrogatif ou relatif

PR 1	qui
PR 2	que
PR 3	quoi
PR 4	qu'
PR 5	lequel
PR 6	laquelle
PR 7	auquel, auxquels auxquelles
PR 8	duquel
PR 9	lesquels, lesquelles
PR 10	desquels, desquelles
PR 11	ou'
PR 12	dont (relatif)

PRONOM (suite)

Indefini

PI 1	on
PI 2	quelqu'un
PI 3	quelques-uns
PI 4	quelques-unes
PI 5	personne
PI 6	rien
PI 7	chacun
PI 8	autre, etc.

Possessif

PO 1	(le, les) mien(s)
PO 2	(le, les) tien(s)
PO 3	(le, les) sien(s)
PO 4	(le, les) no'tre(s)
PO 5	(le, les) vo'tre(s)
PO 6	(la, les) mienne(s)
PO 7	(la, les) tienne(s)
PO 8	(la, les) sienne(s)
PO 9	(la, les) leur(s)

ADJECTIFS

Qualificatif: AQ

Numeral cardinal: AC

Numeral ordinal: AN

Demonstratif: AD
AD 1 ce
AD 2 cet
AD 3 cette
AD 4 ces

Interrogatif: AT
AT 1 quel(s)
AT 2 quelle(s)

ADJECTIFS (suite)

Possessif: AO
AO 1 mon
AO 2 ton
AO 3 son
AO 4 ma
AO 5 ta
AO 6 sa
AO 7 mes
AO 8 tes
AO 9 ses
AO 10 nos
AO 11 vos
AO 12 notre
AO 13 votre
AO 14 leur(s)

Indefini
AI autre, etc.

NOM COMMUN: NC

NOM PROPRE: NP

ADVERBES: DV

DV 0 inde'termine'
DV 1 temps
DV 2 lieu
DV 3 quantite'
DV 4 manie're
DV 5 affirmation
DV 6 ne'gation
DV 7 interrogation

PREPOSITIONS: EP

EP 0 indetermine
EP 1 de, d'
EP 2 du, d'
EP 3 des, d'

EP 4	a'
EP 5	au, aux
EP 6	de, d' (avec verbe)
EP 7	a' (avec verbe)

CONJONCTIONS DE SUBORDINATION: CS

CONJONCTIONS DE COORDINATION: CC

INTERJECTIONS ET ONOMATOPEES: KO

LOCUTION: LC

LC 0	inde'termine'
LC 1	locution adverbiale
LC 2	locution pre'positive
LC 3	locution interrogative

SANS SIGNIFICATION: NS

TABLE 5
Verb Coding Scheme

Basic Tense/Mood	Coding	Person/Number
Infinitive and similar phonetic forms	0	
Present indicative	1,2	basic singular forms
	3	1st person plural
	4	2nd person plural
	5	3rd person plural
Imperfect	6	persons sing. and 3rd plural
	7	1st person plural
	8	2nd person plural

Future	9	1st person singular
	10	2nd and 3rd persons sing.
	11	1st and 3rd persons plural
Present participle	12	
Conditional	13	1st person plural
	14	2nd person plural
Subjunctive	15	three persons sing.
	16	1st person plural
	17	2nd person plural
Simple past	18	three persons singular
	19	3rd person plural
Past participle	20	
Imperative	21	(when followed by a pronoun)

TABLE 6
Examples of Coding for Verbs

Tense/Mood	Coding	Sample French Verbs				
Infinitif	0	arriver	finir	pouvoir	aller	etre
indicatif	1	arrive	finis	peux	vais	suis
	2				vas	es
	3	arrivons	finissons	pouvons	allons	sommes
	4		finissez	pouvez		etes
	5		finissent	peuvent	vont	sont
imparfait de l'indic.	6	arrivais	finissais	pouvais	allais	etais
	7	arrivions	finissions	pouvions	allions	etions
	8	arriviez	finissiez	pouviez	alliez	etiez
futur de l'indic.	9	arriverai	finirai	pourrai	irai	sera
	10	arriveras	finiras	pourras	iras	seras
	11	arriverons	finirez	pourrez	irez	serez
participe present	12	arrivant	finissant	pouvant	allant	etant

conditionnel present	13	arriverions	finirions	pourrions	irions	serions
	14	arriveriez	finiriez	pourriez	iriez	seriez
subjonctif present	15				aille	sois
	16					soyons
	17					soyez
passe simple de l'indic.	18					fus
	19	arriverent	finirent	purent	allerent	furent
participe passe	20			pu		ete
imperatif	21		finis-le			

Articles. Under this general category we have assigned the indefinite articles (*un*,*une*) the category IN, and the definite articles *le*,*la*,*l'*, *les* the category DN. Even though *un* and *une* can be considered as cardinal adjectives, in the present corpus it seemed appropriate to classify them unambiguously as articles.

The words *du*,*de*,*des*,*d'* have been coded as prepositions in the classification EP, which also includes *au* and *aux*. This follows a suggestion made in Dubois (1965, p. 152).

Nouns. Two main subcategories of nouns have been distinguished as in traditional grammars, namely, common and proper nouns. As already indicated, common nouns have been grammatically coded in such a way as to take into consideration the gender and the number. For example, *bouteille* and *bouteilles* have the same spoken form and both have been coded NC2.0, with the first digit indicating the gender (in this case feminine) and the second digit indicating the number, (in this case undetermined). In contrast, *cheval* was coded NC1.1 and *chevaux* was coded NC1.2. Homonyms were coded 0 for both the number and the gender; for example, *le moule* and *la moule* were both coded NC0.0.

In addition, nouns followed by a hyphen and *la'*, which indicates that the noun had the composed form of the demonstrative adjective as determinant, were coded with the symbol # indicating that two words were conjoined. For example, the word *aiguille-la'* was coded NC2.0#AD. Thus NC2.0 is the appropriate coding for *aiguille* and AD the appropriate coding for *la'*.

Proper nouns have been coded to reflect gender only, for example, *Jeanine* was coded NP2 *Papa* NP1, and *Paris* NP0.

Pronouns. Five subcategories have been distinguished: personal (PE), demonstrative (PD), interrogative and relative (PR), indefinite (PI), and possessive (PO). Personal (PE) and demonstrative pronouns (PD) have been given specific numbers, as shown in Table 4.

Interrogative and relative pronouns have been coded together because they are not distinguishable. An exception is *dont*, which is always a relative. Some indefinite pronouns (PI), which could be identified as such unequivocally, have been assigned a specific number from 1 to 7. The other forms which could also be classified as indefinite adjectives have been coded PI8. Because the possessive pronouns are formed with the definite article and the accentuated (*tonique*) form of the possessive adjectives, they do not appear as separate entries in the vocabulary. On the other hand, since the accentuated forms are rarely used as possessive adjectives, they have been categorized as possessive pronouns in Table 4.

Adjectives. Six main subcategories of adjectives have been distinguished: qualitative (AQ), numerical (AC for cardinal and AN for ordinal), demonstrative (AD), interrogative (AT), possessive (AO), and indefinite adjectives (AI). Relative adjectives have not been coded, since their usage is extremely rare even in written language.

Qualitative adjectives have been coded according to the gender, the number and the position (1 for anteposition, 2 for postposition, and 0 for undetermined). For example, *verte* has been coded AQ2.0.2, with the first digit indicating the gender, the second the number and the third the position. In the case of gender and number, the same three-fold classification was used as described above, namely, 0 for indeterminate, 1 for masculine, and 2 for feminine, with corresponding conventions for numbers.

Cardinal adjectives (such as *deux*) have been coded AC. As previously mentioned, the indefinite articles *un* and *une* could be classified as cardinal adjectives, but we classified them IN instead. Ordinal adjectives have been coded according to the gender and number; for instance, *second* has been coded AN1.0. Demonstrative, interrogative, and possessive adjectives have been coded with specific numbers, as shown in Table 1.

Verbs. Although the most elaborate set of distinctions is required for verbs, we especially wanted to reduce the large number of written forms to the much smaller number of spoken forms. A first distinction is among transitive (VT), intransitive (VI), auxiliary and semiauxiliary (VA), impersonal (VM) and pronominal (VR) verbs. The pronominal verbs include reflexive verbs and what Gre'visse calls *verbes pronominaux subjectif*. An example of a reflexive verb is *se regarder*; subjective examples are *se souvenir* and *se douter*. In these latter cases the pronoun has no real semantic role. A sixth entry (VN) was used for verbs which are either transitive or intransitive. Auxiliary and semiauxiliary verbs, which can be also transitive or intransitive, were not coded VT but rather were coded as VI,VA or VT,VA. For example, *ai* as in *j'ai* was coded VT1,VA1 and *es* as in *tu es* was coded

VI2,VA2. In assigning the verbs to each of these six groups we used extensively the French dictionary *Le petit Robert* (1970).

Getting the coding to reflect the mode, tense, and person of the inflectional patterns of spoken French was the most difficult aspect of our task. Before reading the more detailed description that follows, the reader may perhaps find it useful to scan the examples in Table 6.

The infinitive form was coded 0, indicating its indeterminate status. When the imperative (second person plural), the past participle or the second person plural of the present indicative differ phonetically from the infinitive, they were also coded 0. For example *arre'ter*, *arre'tez*, and *arre'te'* are all coded VN0[1]. Concerning the rest of the present indicative, the three singular forms as well as the third person plural were coded 1 when they do not differ phonetically from each other. For example (*je*) *mange*, (*tu*) *manges*, (*ils*) *mangent* were all coded VT1. When the second person and the third person singular differ from the first person singular, they were coded 2. For example (*j'*) *ai* was coded VT1,VA1; (*tu*) *as* and (*il*) *a* were coded VT2,VA2. The first person plural (ending *ons*) was coded 3, as in (*nous*) *voyons*, VN3. The second person plural (ending

[1] The examples given for the coding of verbs are not necessarily extracted from the corpus, since several forms were not used by Philippe.

ez) was coded 4 when it differs phonetically from the infinitive; for example (*vous*) *finissez* was coded VN4; (*vous*) *chantez* was coded VNO. When the third person plural differs phonetically from the three singular forms it was coded 5; for example (*ils*) *chantent* was coded VN1, (*ils*) *servent* was coded VT5, and (*ils*) *vont* was coded VI5.

Concerning the imperfect (*imparfait de l'indicative*), the three singular forms and the third person plural do not differ phonetically from each other. They were all coded 6; for example, (*je*) *fermais*, (*tu*) *croyais*, (*il*) *tenait* and (*ils*) *montaient* were coded VN6. The first and second persons plural (ending with *ions* and *iez*, respectively) were coded 7 and 8, when they differ phonetically from the first and second person plural of the present indicative or from the infinitive. For example, (*nous*) *riions* was coded VI3; (*nous*) *trouvions* was coded VT7; (*vous vous*) *me'fiiez* was coded VIO; and (*vous*) *arriviez* was coded VI8.

For the future indicative, the first person singular (ending *rai*) and the second person plural (ending *rez*) were coded 9; for example, (*j'*) *attraperai* and (*vous*) *arriverez* were coded VT9 and VI9 respectively. The second and the third person singular (endings *ras* and *ra*) were coded 10; (*tu*) *partiras* was coded VI10, (*il*) *sortira* VN10. The first and third person plural (endings *rons* and *ront*) were both coded 11; for example, (*nous*) *danserons* was coded VN11, (*ils*) *de'moliront* VT11.

The present participle was coded 12; for example, *enlevant* was coded VT12.

The three singular forms (endings *rais*, *rais* *rait*) and the third person plural (ending *raient*) of the conditional present (*conditionnel pre'sent*) were difficult to distinguish from the first person singular of the future, in Philippe's speech. Consequently, they have all been coded 9. For example, (*il*) *soignerait* and (*ils*) *mangeraient* were coded VT9. The first and second person plural (endings *rions* and *riez*) were coded 13 and 14, respectively; for example, (*nous*) *recevrions*, VT13, and (*vous*) *prendriez*, VN14.

The simple past (*passe' simple*) is rarely used in spoken language and, as we anticipated, it was not used by Philippe, at least not in the first and the second persons plural. The three singular forms have been coded 18 when they differ phonetically from the other verbal forms which had already been assigned a lower number; for example, (*tu*) *mangeas* was coded VT18, (*il*) *prit* was coded VN18. Verbs for which the first person singular ended with *ai* were difficult to differentiate from the first person singular of the imperfect. Consequently, they received the same code as the first person of that tense, namely, 6; for example, (_je_) _mangeai_ was coded VT6. Verbs for which the three singular forms of the simple past are not different from the first person present indicative have been coded 1; for example, (_tu_) _finis_ has been coded VN1. Finally, the third person plural of the simple past has been coded 19; for example, (_ils_) _dormirent_ has been coded VI19, and (_ils_) _rec*urent_ has been coded VT19.

We turn now to the present subjunctive (_subjonctif pre'sent_). For many verbs, the three singular forms of the present subjunctive do not differ phonetically from some persons of the present indicative. When the three singular forms and the third person plural of the subjunctive are similar (phonetically) to the first person singular of the present indicative, they were coded 1. When they are similar to the third person plural of the present indicative, they were coded 5. For example, (_que je_) _chante_ was coded VN1, and (_qu'il_) _finisse_ was coded VN5. Verbs with a specific form for the first, second and third persons singular were coded 15; for example, (_que j'_) _aille_, (_que tu_) _ailles_, (_qu'il_) _aille_ were all coded VI15.

The first and second person plural of the present subjunctive are in many cases similar to the first and second person plural of the imperfect indicative. Consequently, they were coded 7 and 8, respectively. For example, (_que nous_) _re'ussissions_ was coded VN7, and (_que_ *vous*) *ramassiez* was coded VT8. However, when the first and second persons plural differ from the imperfect, they were coded 16 and 17; for example, (*que nous*) *ayons* was coded VT16,VA16 and (*que vous*) *ayez* was coded VT17,VA17.

Concerning the past participle, when it is not different from the infinitive, it was coded 0. For example, *ferme'* was coded VN0. When the past participle is not different phonetically from the first person of the present indicative, it was coded 1. For example, *fini* was coded VN1, *dit* was coded VT1. When the past participle is different from both infinitive and present indicative, it was coded 20, even when it does not differ phonetically from the simple past. In this one instance we did not follow the basic rule of always coding with the smallest number since the simple past is less frequently used by Philippe than the past participle. Consequently, *recu* and *partis*, which do not differ from (*je*) *recus* and (*je*) *partis*, respectively, were coded VT20 and VI20.

Finally, we turn to the imperative. The imperative does not have endings differing from the indicative present; consequently, it could not be separately categorized when there was no reference to the context. On the other hand, in written language a hyphen is used to link the verb to the personal pronoun; this same practice was followed in transcribing the corpus, which means that in the transcribed corpus, the imperative-personal pronoun can be identified as a unit. In these cases the imperative form was coded 21, followed by the appropriate code for the personal pronoun; for example, *donne-moi* was coded VN21#PE3. Only the second person singular of the imperative, followed by a pronoun, occurred in Philippe's speech, so the first and second person plural of the imperative were not coded.

One or two general aspects of the coding of verbs need to be emphasized. In the first place, we have coded with the smallest number possible for a given form. Thus, the past participle *allé* or the imperative *allez* has been coded 0, not a higher number. In other words, the rule followed was to code with the lowest number consistent with the phonetic form, except when the past participle had the same form as the simple past. It is also important to emphasize that the individual vocabulary words were coded independently of their context and, consequently, composed forms, e.g., *j'aurai rec*u*, were not coded as a single verb form but rather the individual words *aurai* and *rec*u* were coded separately.

Finally, tenses that are extremely infrequent in spoken speech have not been coded, such as the imperfect subjunctive (*imparfait du subjonctif*).

IV. NOUN-PHRASE GRAMMAR I

Our study of the noun phrases of Philippe's speech is based on a selection of noun phrases made from the sentences in 9 of the 33 sessions of the complete corpus. For this selection, we used Sessions 1, 2, 3, 14, 15, 16, 31, 32 and 33, which cover the beginning, the middle and the end of the corpus. In other words, each of these sets of three sessions forms a *section* of the corpus. This choice was motivated primarily by a desire to look for the changes in structure over the maximum period of time.

The noun phrases were selected by hand from the sentences in the sessions using interactive computer programs. The selection of the noun phrases was completely the judgment of M. Le'veille', who is a native French speaker, and was not the product of any formal analysis. Nevertheless certain guidelines were followed and they are listed below.

1. Noun phrases in questions of the forms *est-ce que*? and *est-ce qu'il*? were not included, although other kinds of interrogatives were scanned for noun phrases.

2. Personal pronouns combined with verbs to form pronominal verbs were not included. An example of this construction is *je m'en vais*; where the personal pronoun *m'* was not included.

3. Cases of using the preposition *a'*, where *de*

should normally have been used (e.g., as in je veux la montre a' maman), as well as phrases such as sa maison a' lui, were included.

4. When the preposition en was combined with a substantive, such as in la chaise en me'tal, the whole expression en me'tal was excluded. On the other hand, when en could be replaced by sur le or dans le, as in en ve'lo or en voiture, it was included.

5. The preposition pour was analyzed in such sentences as un bassin pour les bateaux.

6. When adverbs such as beaucoup or assez modified a substantive, as in beaucoup de chocolat, assez de sous, they were included in the noun phrases.

7. When the word la' followed a substantive, such as in appuie sur le bouton la', the word was considered a part of the noun phrase.

8. When a word could serve as either a noun or another word, such as te'le'phone, the word was not included as a noun phrase if the context of usage was vague.

We expect the class of noun phrases to change somewhat in a complete grammar, and the rules of the two grammars discussed below to change in a full grammar. For example, one motivation for change will be the attempt to study differences between noun phrases that occur as subjects in some sentences and as direct objects in others. As for the methodological justification for selecting the noun phrases by hand, we should mention that this was only a first step in obtaining a complete grammar. The fact that noun phrases dominated Philippe's speech, as they dominate other psycholinguistic corpora, indicates that the noun phrase is a place to begin.

It may be useful to expand on these methodological remarks about our method of selecting the noun phrases by individual inspection of the utterances. In principle, one would like to put the entire grammar and semantics together into one systematic whole. We think this is a worthy objective and hope to pursue it further in subsequent publications. On the other hand, given the massive character of a corpus of the size of this one, we consider it important to begin the analysis in piecemeal stages, and to do this we must make some intuitive decisions without systematic guidance from an overall model of a complete grammar or even less a complete semantics. We are certain that there could be marginal disagreements about our selection of noun phrases but we doubt that it would affect the main results of our analysis. We emphasize also that the noun phrases were selected prior to the attempt to write a grammar for them, and consequently we have in no sense tailored the selection of noun phrases to the convenience of our two grammars.

As already indicated, there are two separate and parallel analyses of the noun phrases in the corpus. In this section we describe the first grammar, the one that uses the first dictionary of grammatical categories. This grammar represents an extensive modification of those used in Suppes (1971, 1972) and Smith (1972) for analyzing noun phrases in English. Grammar I is of course not a grammar of English, but rather it was constructed from the modifications that would be required in the grammars for English in order to obtain an appropriate grammar for French. Grammar II, on the other hand, was written by a more direct consideration of French grammars. Detailed remarks about the goodness of fit are reserved for later discussion, after both grammars have been described and the results of the analysis presented.

Nonterminal vocabulary of Grammar I. In addition to the categories introduced in Dictionary I, the following additional nonterminal symbols were used in Grammar I. These additional symbols cannot be rewritten by lexical rules, but occur in derivation trees always at least two nodes from a terminal node. As the label of the root of all the noun-phrase trees, we adopted SN for the French *syntagme nominal*. Two additional noun-phrase notations for which we used a natural English notation, namely, NP* and NP', were also introduced. Note that according to Dictionary I, NP without an additional suffix was the category of proper names.

In the production rules of the grammar as shown in Table 7, we omitted all the numerical suffixes showing number, case, gender, etc. These suffixes were included in the dictionaries, but were ignored by the grammar discussed here. In a more refined analysis it would be necessary to add these subscripts. However, it is obvious that the refinement required to do this would multiply the number of rules. While a relatively abstract level of analysis has been chosen, we feel the level selected exemplifies the structural features of Philippe's speech at about the right level of conceptual detail, insofar as noun phrases are concerned.

TABLE 7

Production Rules of Grammar I

(1,1) SN -> NP* CC NP*
(1,2) SN -> NP* EP NP*
(1,3) SN -> NP*
(1,4) SN -> EP NP*
(1,5) SN -> CC NP*
(1,6) SN -> KO NP*
(1,7) SN -> NP* KO
(1,8) SN -> LC NP*
(1,9) SN -> NP* LC

(2,1) NP' -> PE
(2,2) NP' -> PD
(2,3) NP' -> PR
(2,4) NP' -> PI
(2,5) NP' -> NP
(2,6) NP' -> NC
(2,7) NP' -> PO

(3,1) NP* -> DET NUM ADJP NP' POSTADJP
(3,2) NP* -> NUM ADJP NP' POSTADJP
(3,3) NP* -> DET ADJP NP' POSTADJP
(3,4) NP* -> ADJP NP' POSTADJP
(3,5) NP* -> DET NUM NP' POSTADJP
(3,6) NP* -> NUM NP' POSTADJP
(3,7) NP* -> DET NP' POSTADJP
(3,8) NP* -> NP' POSTADJP
(3,9) NP* -> DET NUM ADJP NP'
(3,10) NP* -> NUM ADJP NP'
(3,11) NP* -> DET ADJP NP'
(3,12) NP* -> ADJP NP'
(3,13) NP* -> DET NUM NP'
(3,14) NP* -> NUM NP'
(3,15) NP* -> DET NP'
(3,16) NP* -> NP'

(4,1) ADJP -> AQ
(4,2) ADJP -> ADJP AQ

(5,1) POSTADJP -> AQ
(5,2) POSTADJP -> POSTADJP AQ

(6,1) DET -> AD
(6,2) DET -> AT
(6,3) DET -> AO
(6,4) DET -> DN
(6,5) DET -> IN
(6,6) DET -> AI

(7,1) NUM -> AC
(7,2) NUM -> AN

The notations ADJP and POSTADJ were introduced for adjectival phrases in pre-position and for adjectival phrases following the noun. The nonterminal symbol DET is the general nonterminal symbol used for determiners. Rules (5,1) to (5,6) are rewrite rules for this nonterminal symbol. The nonterminal symbol NUM designates numerical or cardinal number words (Rules (6,1) and (6,2)).

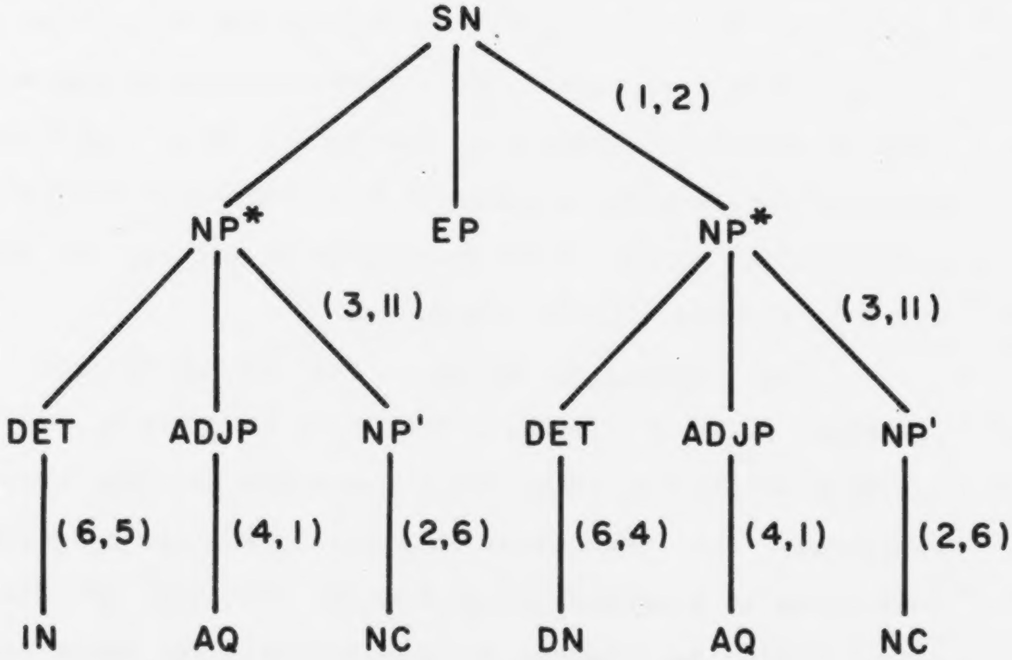
We have already mentioned that we introduced two noun-phrase symbols in addition to SN, namely, NP′ and NP*. The function of NP′ was to provide a rewrite rule for one-word noun phrases (Rules (2,1) to (2,7)). Rewrite rules (3,1) to (3,16), that have NP* on the left, were used to construct noun phrases of a more complicated character. From a purely grammatical standpoint both nonterminal symbols were needed in order to block improper recursions. Further, both symbols were convenient in getting a better fit from a probabilistic standpoint.

The production rules. The 44 production (or rewrite) rules of Grammar I are shown in Table 7. Perhaps the most surprising thing about the rules is their standard character. At the level of syntax, Philippe's speech is very close to standard spoken French. Not one of the 44 rules would be regarded as inappropriate for adult spoken French; but additional rules beyond those given here would be needed in a complete grammar.

The use of the 44 production rules to produce an infinite variety of noun phrases is standard and in accordance with classical context-free grammars. An illustration of the use of the rules to produce one of the longer noun phrases in Philippe's speech is shown in the tree of Figure 1 for IN AQ NC EP DN AQ NC.

Figure I

TREE FOR IN AQ NC EP DN AQ NC


SN
(1,2)
NP*
EP
NP*
(3,11)
(3,11)
DET
ADJP
NP'
DET
ADJP
NP'
(6,5)
(4,1)
(2,6)
(6,4)
(4,1)
(2,6)
IN
AQ
NC
DN
AQ
NC

There are some additional conceptual remarks that we want to make about the rules. The first group of rules constitutes the nine different rules that can be used to generate trees starting from the labeled root SN. The first three rules are standard. Rule (1,4), SN -> EP NP*, a rule which permits a noun phrase to begin with a preposition, is natural not only for Philippe's speech, but also for standard spoken French. The instances of the use of this rule in accounting for noun phrases in the corpus are the following:

Frequency	Noun phrase
244	EP NC
60	EP DN NC
11	EP AQ NC
11	EP NP
6	EP NC AQ
3	EP AI NC
3	EP AT NC
2	EP IN NC
2	EP PI
1	EP AO NC
1	EP AQ AQ NC
1	EP AQ NC AQ
1	EP DN AQ NC
1	EP PR

Types = 14 Tokens = 347

Rule (1,5), SN -> CC NP*, introduced conjunctions at the beginning of noun phrases. This rule, like Rule (1,4), is natural for adults as well as for Philippe. The instances of Rule (1,5) in the corpus are the following:

Frequency	Noun phrase
11	CC PD
4	CC PE
1	CC DN NC
1	CC PE AQ

Types = 4 Tokens = 17

Rules (1,6) and (1,7) introduced the grammatical category KO (interjections) before and after the main noun phrases, respectively. Rule (1,6) was not required for our noun-phrase fragment. The instances of Rule (1,7) are the following:

Frequency	Noun phrase
8	PD KO
4	AD NC KO
3	AO NC KO
1	AD AQ NC KO
1	DN PO KO

Types = 5 Tokens = 17

In a similar way, Rules (1,8) and (1,9) introduced the grammatical category LC (locutions) before and after noun phrases. Rule (1,8) was not required for the noun-phrase fragment. The only instance of (1,9) was the phrase PE LC, which occurred once in the corpus.

As already remarked, the rules of the second group were used to generate one-word noun phrases. These rules were generated by using Rules (1,3) and (3,16), and then rewriting NP' according to one of the seven rules in the second group.

The 16 rules of the third group contain the main structure of the noun phrases. They determine the way in which determiners, numerical expressions, preadjectival phrases, and postadjectival phrases can be generated. Although all 16 rules are perhaps plausible to a native French speaker, three of the rules from the third group were not required for the corpus. The unused rules from the third group were (3,1), (3,2), and (3,5). An instance of Rule (3,3) is[2]

2 DN AQ NC AQ

and an instance of Rule (3,9) is

2 DN AC AQ NC

The fourth and fifth groups of rules—(4,1), (4,2), (5,1), (5,2)— provided the standard recursions for pre- and post-adjectival phrases. Except for accounting for both pre- and post-positions, they were exactly the same rules used previously in English.

The main reason for using different rules for pre- and post-position adjectival phrases was to test the tendency to use a different number of adjectives in the pre-position from the postposition. After we examine the fit for the grammar in more detail, we will remark on the outcome of this test.

The sixth group of rules is for determiners and permitted rewriting the nonterminal symbol DET as any one of six nonterminal symbols standing for a grammatical category. Note that determiners were not introduced as a separate grammatical category, but were spread among the grammatical categories of Dictionary I, which is a relatively classical structure. In a sense, this was a way of having the best of both worlds. We introduced the concept of a determiner in the grammar, but kept in the dictionary of grammatical categories the more or less classical divisions.

---

[2] Here and subsequently we often show the frequency of a given grammatical type of noun phrase in the corpus without expressly labeling the frequency column. Thus the frequency of DN AQ NC AQ is 2 in the noun phrases under examination.

Finally, the seventh group of rules consists of the two simple rules for rewriting the nonterminal symbol NUM for numerical expressions, which may be replaced with the grammatical categories AC and AN. Rule (7,2), which introduces AN, was not used in the corpus.

Nonprobabilistic fit of Grammar I. Probably the first empirical question to ask about a grammar for a corpus is, what percentage of the tokens and what percentage of the grammatical types does it parse? The data are summarized in Table 8. There were 5,547 noun phrases in the nine hours analyzed. Grammar I parsed 5,266 of these noun phrases, which represents 94.9 percent of the tokens. There were 345 types of noun phrases in the nine hours, and the grammar parsed 223 of these, or 64.6 percent.

TABLE 8

Nonprobabilistic Fit of Grammar I Noun-phrase Corpus

Description	Types	Tokens
Total	345	5547
Recognized by Grammar I	223	5266
Percent recognized	64.6%	94.9%
Lexically ambiguous	244	1830
Percent lexically ambiguous	70.7%	33%
Ambiguous recognized	168	1716
Percent of the ambiguous recognized	68.9%	93.8%
Ambiguity resolved	149	1666
Ambiguity reduced	17	43

Original ambiguity factor = 2806

Parsed ambiguity factor = 2480

Reduced ambiguity factor = 52

Types after consolidation = 143

The entries in Table 8 are defined as follows, with the usual distinction between types and tokens. We first give the total number of types and tokens in the corpus, we second give the number recognized, and third the percent recognized. The fourth line indicates how many of the types or tokens in the corpus are lexically ambiguous, that is, have at least one word assigned to more than one grammatical category. It is important to realize that lexical ambiguity, as we have defined it, results from using words in ways that are grammatically distinct; in most cases, lexical ambiguity would offer no difficulty to a speaker. It is, we believe, a test of the grammar to ask how well it is able to dispense with this superficial ambiguity.

We then indicate how many of these lexically ambiguous types or tokens are recognized by the grammar. When the grammar recognizes a lexically ambiguous type, it is important to realize that the type is a 'shorthand' notation for several types.

For example:

(1) DN,PE NC

occurs 732 times in the original corpus. The first word, according to Dictionary I, could be either DN or PE, and hence (1) expands to

(1′) DN NC
(1″) PE NC .

In this case (1′) was recognized by Grammar I while (1″) was not; hence, we say that the lexical ambiguity was resolved, meaning that only one of the alternative types was recognized. Another lexically ambiguous case is:

(2) 7 DN,PE AQ,PI,AI ,

which represents the types

(2.1) DN AQ
(2.2) PE AQ
(2.3) DN PI
(2.4) PE PI
(2.5) DN AI
(2.6) PE AI .

In this case, (2.2) and (2.3) were both recognized, and so we say that the lexical ambiguity was *reduced* (but not resolved). This reduction happened on 17 types (43 tokens) in the corpus. Nineteen lexically ambiguous types (50 tokens) remained that were not resolved by Grammar I. The elimination of this remaining lexical ambiguity will be discussed further after we have discussed the probabilistic fit of Grammar I.

The *ambiguity factor* is, intuitively, the number of 'extra' analyses of a noun phrase. Let S be the set of types in a corpus, for t in S, f(t) is the *frequency* of t, n(t) is the *number* of types (using simple symbols) that arise from t. Then, the *original ambiguity factor* is given by

(3) $$\sum_{t \text{ in } S} f(t)*(n(t)-1) .$$

We subtract 1 in order to eliminate an entry for those types that have a single classification. For example, the contribution of

9 DN,PE NC,VN

to the original ambiguity factor is

$9*(4-1)=27$,

where 9 is the frequency, and 4 is the number of simple types arising from the type in question.

The parsed ambiguity factor is defined in a similar fashion, except now we use S' instead of S, where S' is the set of types in S' recognized by the grammar.

The reduced ambiguity factor is given by the formula

$$\sum_{t \text{ in } S'} f(t)*(n'(t)-1) ,$$

where S' is the subset of S recognized by the grammar, and n'(t) is the number of simple types arising from t that are recognized by the grammar. Among the measures we consider, the reduced ambiguity factor is clearly the best measure of the lexical ambiguity permitted by the grammar. As will be seen, this number is quite small for all of the analyses presented in this work.

The resolution and reduction of lexical ambiguity alters the list of types present in the corpus. For example, 732 DN,PE NC

is resolved to

732 DN NC ,

(since PE NC is not recognized by Grammar I) and must be combined with

34 DN,PE NC,VT ,

which is also resolved to DN NC. This process is called consolidation. Of 223 original types recognized by Grammar I, 143 types remained after consolidation.

Probabilistic fit of Grammar I. Following the standard notion of a probabilistic grammar and the various ways of estimating the goodness of fit (Suppes, 1970, 1971; Smith, 1972), we used a model based on a geometric distribution for predicting the parameters of the grammars reported upon here.

Table 9 gives the parameters for the standard fit obtained by the maximum-likelihood method, and Table 10 shows the results of the goodness-of-fit test for each noun-phrase type. Types with expected frequency less than 5 were grouped together to avoid the problem of enormous chi-square contributions. The residual was a relatively high 1112.15, which indicates that over 20 percent of the noun phrases predicted by the grammar did not occur in the corpus. There were 13 groups (low-expected types clumped together), thus reducing the degrees of freedom to 12. The chi-square was 5527, and the chi-square divided by the degrees of freedom was 460.62.

TABLE 9

Parameters for Grammar I*
Philippe's Noun Phrases

Label	Prob.	Rule
(1,1)	.0074	SN -> NP* CC NP*
(1,2)	.0243	SN -> NP* EP NP*
(1,3)	.8938	SN -> NP*
(1,4)	.0659	SN -> EP NP*
(1,5)	.0032	SN -> CC NP*
(1,6)	.0000	SN -> KO NP*
(1,7)	.0051	SN -> NP* KO
(1,8)	.0000	SN -> LC NP*
(1,9)	.0002	SN -> NP* LC
(2,1)	.2716	NP' -> PE
(2,2)	.1596	NP' -> PD
(2,3)	.0449	NP' -> PR
(2,4)	.0337	NP' -> PI
(2,5)	.0731	NP' -> NP
(2,6)	.4143	NP' -> NC
(2,7)	.0028	NP' -> PO
(3,1)	.0000	NP* -> DET NUM ADJP NP' POSTADJP
(3,2)	.0000	NP* -> NUM ADJP NP' POSTADJP
(3,3)	.0007	NP* -> DET ADJP NP' POSTADJP
(3,4)	.0002	NP* -> ADJP NP' POSTADJP
(3,5)	.0000	NP* -> DET NUM NP' POSTADJP
(3,6)	.0002	NP* -> NUM NP' POSTADJP
(3,7)	.0064	NP* -> DET NP' POSTADJP
(3,8)	.0092	NP* -> NP' POSTADJP
(3,9)	.0004	NP* -> DET NUM ADJP NP'
(3,10)	.0006	NP* -> NUM ADJP NP'
(3,11)	.0271	NP* -> DET ADJP NP'
(3,12)	.0127	NP* -> ADJP NP'
(3,13)	.0006	NP* -> DET NUM NP'
(3,14)	.0037	NP* -> NUM NP'
(3,15)	.2790	NP* -> DET NP'
(3,16)	.6594	NP* -> NP'
(4,1)	.9700	ADJP -> AQ
(4,2)	.0300	ADJP -> ADJP AQ
(5,1)	.9891	POSTADJP -> AQ

(5,2)	.0109	POSTADJP -> POSTADJP AQ
(6,1)	.0334	DET -> AD
(6,2)	.0070	DET -> AT
(6,3)	.0756	DET -> AO
(6,4)	.6225	DET -> DN
(6,5)	.2515	DET -> IN
(6,6)	.0100	DET -> AI
(7,1)	1.0000	NUM -> AC
(7,2)	.0000	NUM -> AN

*Correction for continuity = .5; expected < 5 are grouped together.

TABLE 10
Noun-phrase Types for Grammar I

Observed	Expected	Chi-square	Count	Noun phrase type
1428	842.551	406.106	1	PE
830	495.249	225.591	1	PD
793	338.487	608.973	1	DN NC
324	226.775	41.256	1	NP
281	136.733	151.163	1	IN NC
244	94.789	233.309	1	EP NC
237	139.378	67.677	1	PR
174	1285.248	959.940	1	NC
156	104.533	24.849	1	PI
107	41.116	103.978	1	AO NC
61	12.869	176.302	1	IN AQ NC
60	24.964	47.779	1	EP DN NC
55	31.856	16.095	1	DN AQ NC
45	24.021	17.460	1	AQ NC
37	18.167	18.500	1	AD NC
18	7.178	14.842	1	AC NC
16	7.735	7.796	1	DN NC AQ
*******				
23			1	IN NC EP NC
21			1	DN NC EP NP
13			1	DN NC EP NC
13			1	IN NC AQ
**GROUP 70	7.100	548.421		
13	11.636	.064	1	PE AQ
12	27.530	8.206	1	DN PI
11	5.418	4.766	1	AI NC

```
*******
          12                           1   DN PO
          11                           1   CC PD
          11                           1   EP AQ NC
**GROUP  34       5.818      131.725
          11     16.725        1.632   1   EP NP
          10     11.121         .035   1   IN PI
          10      9.551         .000   1   NC EP NC
*******
          10                           1   DN NC KO
           9                           1   PE AQ CC PE AQ
           8                           1   AD AQ NC
           8                           1   AT NC
**GROUP  35       7.477       97.662
           8     17.750        4.821   1   NC AQ
           5      9.256        1.524   1   AQ PD
           5     59.724       49.230   1   DN NP
*******
           8                           1   NP CC NP
           8                           1   PD KO
           7                           1   DN NC EP DN NC
           6                           1   EP NC AQ
           5                           1   IN AQ NC EP NC
           4                           1   AD NC KO
**GROUP  38       5.104      205.618
*******
           4                           1   AO AQ NC
           4                           1   AO NC EP DN NC
           4                           1   CC PE
**GROUP  12       6.994        2.903
           3      7.255        1.943   1   AO NP
*******
           4                           1   DN NC CC DN NC
           4                           1   IN NC CC IN NC
           3                           1   AC AQ NC
           3                           1   AI AQ NC
           3                           1   AO NC KO
           3                           1   AO NC EP PE
           3                           1   DN NC EP PE
           3                           1   DN NC EP AD NC
           3                           1   DN NC EP IN NC
           3                           1   EP AI NC
           3                           1   EP AT NC
           3                           1   NC EP NP
**GROUP  38       6.545      146.419
*******
           3                           1   PD CC PD
           2                           1   AD NP
           2                           1   AQ AQ NC
           2                           1   DN AC AQ NC
```

```
       2                               1   DN AQ NC AQ
**GROUP  11     5.650         4.163
       2        36.707        31.877   1   DN PR
       2        10.084         5.704   1   EP IN NC
       2         7.709         3.520   1   EP PI
*******
       2                               1   DN NC CC DN NC AQ
       2                               1   IN AC NC
       2                               1   NC EP DN NC
       2                               1   NP AQ
**GROUP   8     5.923          .420
*******
       2                               1   PD EP NP
       2                               1   PD EP PD
       1                               1   AC NC AQ
       1                               1   AC NC CC AC NC
       1                               1   AD AQ NC KO
       1                               1   AD NC AQ
       1                               1   AO PI
**GROUP   9     6.192          .861
*******
       1                               1   AO PR
       1                               1   AQ NC EP NC
       1                               1   AQ NC EP NP
       1                               1   AQ NC EP IN NC
       1                               1   AQ NC CC DN NC
       1                               1   AT AQ NC
**GROUP   6     5.062          .038
*******
       1                               1   CC DN NC
       1                               1   CC PE AQ
       1                               1   DN AC NC
       1                               1   DN AQ NC EP NP
       1                               1   DN NC EP PR
       1                               1   DN NC EP PD
       1                               1   DN NC EP AO NC
       1                               1   DN NC CC AO NC
       1                               1   DN NC EP AQ NC
       1                               1   DN PO KO
       1                               1   DN PO EP NC
       1                               1   DN PO EP NP
       1                               1   EP AO NC
**GROUP  13     6.436         5.713
       1        10.279         7.498   1   EP PR
*******
       1                               1   EP AQ AQ NC
       1                               1   EP AQ NC AQ
       1                               1   EP DN AQ NC
       1                               1   IN AQ NC AQ
       1                               1   IN AQ AQ NC
       1                                    1   IN AQ AQ AQ NC
       1                                    1   IN AQ AQ NC AQ
       1                                    1   IN AQ NC EP DN NC
       1                                    1   IN AQ NC EP DN AQ NC
       1                                    1   IN NC EP PR
       1                                    1   IN NC EP AD NC
       1                                    1   IN NC EP IN NC
       1                                    1   IN NC EP DN NC
       1                                    1   IN NC EP IN NC AQ
       1                                    1   IN NC EP IN AQ NC
       1                                    1   IN NC EP DN AQ NC
       1                                    1   IN NC AQ CC IN NC AQ
       1                                    1   IN PI EP NC
       1                                    1   NC AQ AQ
       1                                    1   NC AQ EP NP
       1                                    1   NC CC NC
**GROUP  21      6.958        26.358
*******
       1                                    1   NC EP PR
       1                                    1   NP CC NC
       1                                    1   NP CC PE
       1                                    1   NP EP NP
       1                                    1   PE CC PE
       1                                    1   PE CC DN NC
       1                                    1   PE LC
       1                                    1   PI EP IN NC
**GROUP   8      4.198         2.597
              1112.1497    1112.1497        Residual

    5264      5264         5527.4833

Degrees of freedom =  12

Chi-square/degrees of freedom =  460.6236

Groups =  13
```

We now examine these tables, with a regard to what they tell us about Philippe's noun phrases. In Table 9 notice that the 4-rules, which generate pre-position adjectival phrases, have nearly the same parameters as the

5-rules, which generate the post-position adjectival phrases. This means our hypothesis that these would be different in character is rejected. The tendency to use more than one AQ is always small, whether in pre- or postposition. Notice among the 3-rules that

(3,15) NP* -> DET NP'

and

(3,16) NP* -> NP'

have between them 92 percent of the probability. Again this limits the expected length of the noun phrase rather sharply (see Appendix B).

Turning to Table 10, one notices that the predicted values for the high-frequency types are rather different from the observed frequency. The problem arises most with the types

PE
PD
DN NC
EP NC
PR ,

which have far too small an expected frequency, and NC, which has far too large an expected frequency.

Notice that the 2-rules, which introduced NC, PE, etc., into longer phrases by replacement in the 3-rules, also performed double duty in creating, through Rules (1,3), (3,16), the one-word phrases that caused so much difficulty. The grammar needs a different mechanism for introducing these symbols (NC, PE, etc.) into longer noun phrases than the mechanism used for the shorter ones.

The large residual suggests that the grammar generated many types not found in the corpus. The 1-rules and 3-rules are mainly responsible for this. Note that Rule (1,4) (SN -> EP NP*) generated many types not found. Also, while

244	EP NC
60	EP DN NC
11	EP AQ NC

were underpredicted by Grammar I, some other types using (1,4) were overpredicted, such as the following:

11	EP NP
2	EP IN NC
2	EP PI
1	EP PR .

Several types using (1,2) (SN -> NP* EP NP*) were underpredicted, including

23	IN NC EP NC
21	DN NC EP NP
13	DN NC EP DN NC
7	DN NC EP DN NC
5	IN AQ NC EP NC .

An exception was

10	NC EP NC ,

which was predicted almost exactly.

Other problems are hidden within groupings of low-frequency utterances. These include:

Observed Frequency	Expected Frequency	Chisquare	Noun-phrase type
21	.444	906.3	DN NC EP NP
9	.000	302,856	PE AQ CC PE AQ
8	.091	605.9	NP CC NP

Note that DN NC EP NP used Rule (1,2) as a further indication of the problems with that rule. PE AQ CC PE AQ and NP CC NP used Rule (1,1) (SN ->NP* CC NP*). The types using (1,1) have difficulties similar to those of (1,2), in that the more complex of them were unpredicted, while

1	NC CC NC
1	PE CC PE

were overpredicted by Grammar I.

We now turn to the geometric fit for Grammar I. The method consisted of taking the rules in each class, ordering them according to the highest observation of their usage, and predicting this observed usage with a geometric distribution. The assumption is that there is a mechanism for selecting which rule of a given class to use, and that we can predict this mechanism with a geometric distribution. The gain resulting from this simplification is a decrease in the number of parameters in the model.

Table 11 shows the appropriateness of the geometric model. The rules in each class are ordered according to their observed usage. The numbers in the "Observed" column represent the number of times the rule was used in the corpus, weighed, of course, against the frequency of the types. The standard geometric distribution with a single parameter 0 was used to predict the values given in the "Expected" column.[3] The parameter 0 is given below each class of rules. (The distributions are *truncated geometric*, since the tail of the distribution is added to the last element of the list. This practice can be troublesome on a very flat distribution, such as that for the rules in group 2, where the tail contains enough probability to inflate the last value. We have, however, found that the distributions are usually not this flat. We should remark that, when there are only two rules in a group (as in group 4), the truncated geometric simply reduces to the binomial.)

It is clear what mechanism is natural to associate with a truncated geometric distribution. We assume that the production rules for rewriting a nonterminal of a given kind are located in a push-down store. By usage and past experience, the rules are stored from the top down in

[3] By the maximum likelihood estimation, the parameter is the reciprocal of the mean of the distribution. We denote this parameter *theta*.

approximate order of usage, that is, with the most frequent being on top, etc. When a given rule is used the speaker goes to the location, selects the first rule with average probability theta, goes to the second rule with probability (1-theta); upon reaching the second rule then selects it with probability theta or with probability (1-theta) goes on to the next rule, and so forth. It is important to interpret these probabilities as being on the average. Obviously, in a given situation semantic and perceptual factors will determine often uniquely which rule will be selected, although in many cases it will still be undetermined and a probabilistic selection will be made. What is important about the assumption of such a model in the present context is that it is meant to represent averaging results of accessing without taking into account particular semantic or perceptual features of a given occurrence of use.

Table 11 gives the parameters of the grammar based on the geometric distribution for comparison to the parameters of the normal distribution. Table 12 gives the fit of the model, based on the geometric. From a qualitative point of view, the difficulties encountered in fit of the model based on the geometric distribution are similar to the problems in the standard fit, as described above.

TABLE 11
Parameters for Geometric Fit of Grammar I

Label	Observed	Expected	Rule
(1,3)	4705	4481.4675	SN -> NP*
(1,4)	347	666.2033	SN -> EP NP*
(1,2)	128	99.0360	SN -> NP* EP NP*
(1,1)	39	14.7224	SN -> NP* CC NP*
(1,7)	27	2.1886	SN -> NP* KO
(1,5)	17	.3254	SN -> CC NP*
(1,9)	1	.0568	SN -> NP* LC
Theta =		.8513426	
(2,6)	2250	2493.3708	NP' -> NC
(2,1)	1475	1348.6649	NP' -> PE
(2,2)	867	729.4932	NP' -> PD
(2,5)	397	394.5830	NP' -> NP
(2,3)	244	213.4300	NP' -> PR
(2,4)	183	115.4444	NP' -> PI
(2,7)	15	136.0138	NP' -> PO
Theta =		.4590998	
(3,16)	3581	3657.4886	NP* -> NP'
(3,15)	1515	1194.3653	NP* -> DET NP'
(3,11)	147	390.0240	NP* -> DET ADJP NP'
(3,12)	69	127.3637	NP* -> ADJP NP'
(3,8)	50	41.5910	NP* -> NP' POSTADJP
(3,7)	35	13.5817	NP* -> DET NP' POSTADJP
(3,14)	20	4.4351	NP* -> NUM NP'
(3,3)	4	1.4483	NP* -> DET ADJP NP' POSTADJP
(3,10)	3	.4730	NP* -> NUM ADJP NP'
(3,13)	3	.1544	NP* -> DET NUM NP'
(3,9)	2	.0504	NP* -> DET NUM ADJP NP'
(3,4)	1	.0165	NP* -> ADJP NP' POSTADJP
(3,6)	1	.00[illegible]	NP* -> NUM NP' POSTADJP
Theta =		.6734466	
(4,1)	226	226.0000	ADJP -> AQ
(4,2)	7	7.0000	ADJP -> ADJP AQ
Theta =		.9699571	

```
(5,1)       91        91.0000      POSTADJP -> AQ
(5,2)        1         1.0000      POSTADJP -> POSTADJP AQ

Theta =        .9891304

(6,4)     1062      1078.3448      DET -> DN
(6,5)      429       396.7343      DET -> IN
(6,3)      129       145.9627      DET -> AO
(6,1)       57        53.7012      DET -> AD
(6,6)       17        19.7572      DET -> AI
(6,2)       12        11.4998      DET -> AT

Theta =        .6320896

(7,1)       29        29.0000      NUM -> AC

Theta =       1.0000000
```

TABLE 12

Noun Phrases Types for Grammar I
Geometric Fit

Observed	Expected	Chi-square	Count	Noun phrase type
1428	749.459	613.430	1	PE
830	405.382	443.719	1	PD
793	285.998	897.013	1	DN NC
324	219.271	49.544	1	NP
281	105.222	291.978	1	IN NC
244	205.976	6.836	1	EP NC
237	118.604	117.192	1	PR
174	1385.577	1058.553	1	NC
156	64.153	130.069	1	PI
107	38.712	118.701	1	AO NC
61	33.328	22.153	1	IN AQ NC
60	42.516	6.785	1	EP DN NC
55	90.588	13.591	1	DN AQ NC
45	46.800	.036	1	AQ NC
37	14.243	34.782	1	AD NC
*******				
23			1	IN NC EP NC
21			1	DN NC EP NP
18			1	AC NC

```
     16                             1       DN NC AQ
**GROUP  78   5.925    864.600
     13       8.430      1.965      1       PE AQ
     12      13.242       .042      1       DN PI
     12      15.601       .616      1       DN PO
     11       5.240      5.280      1       AI NC
     11       6.957      1.804      1       EP AQ NC
     11      32.596     13.654      1       EP NP
*******
     13                             1       DN NC EP NC
     13                             1       IN NC AQ
     11                             1       CC PD
     10                             1       DN NC KO
     10                             1       IN PI
**GROUP  57   8.179    285.499
     10       9.467       .000      1       NC EP NC
*******
      9                             1       PE AQ CC PE AQ
      8                             1       AD AQ NC
      8                             1       AT NC
**GROUP  25   7.561     37.946
      8      15.585      3.221      1       NC AQ
      5      13.692      4.902      1       AQ PD
      5      45.260     34.928      1       DN NP
      4      12.262      4.913      1       AO AQ NC
*******
      8                             1       NP CC NP
      8                             1       PD KO
      7                             1       DN NC EP DN NC
      6                             1       EP NC AQ
      5                             1       IN AQ NC EP NC
      4                             1       AD NC KO
      4                             1       AO NC EP DN NC
      4                             1       CC PE
      4                             1       DN NC CC DN NC
      4                             1       IN NC CC IN NC
      3                             1       AC AQ NC
      3                             1       AI AQ NC
**GROUP  60   5.199    567.194
      3       6.126      1.126      1       AO NP
*******
      3                             1       AO NC KO
      3                             1       AO NC EP PE
      3                             1       DN NC EP PE
      3                             1       DN NC EP AD NC
      3                             1       DN NC EP IN NC
      3                             1       EP AI NC
      3                             1       EP AT NC
      3                             1       NC EP NP
      3                             1       PD CC PD
      2                                1      AD NP
**GROUP  29   6.492        74.601
      2      24.481        19.736      1      DN PR
      2      15.642        11.042      1      EP IN NC
      2       9.537         5.192      1      EP PI
*******
      2                                1      AQ AQ NC
      2                                1      DN AC AQ NC
      2                                1      DN AQ NC AQ
      2                                1      DN NC CC DN NC AQ
      2                                1      IN AC NC
      2                                1      NC EP DN NC
      2                                1      NP AQ
**GROUP  14   6.185         8.651
*******
      2                                1      PD EP NP
      2                                1      PD EP PD
      1                                1      AC NC AQ
      1                                1      AC NC CC AC NC
      1                                1      AD AQ NC KO
      1                                1      AD NC AQ
      1                                1      AO PI
      1                                1      AO PR
**GROUP  10   6.520         1.362
      1       5.755         3.146      1      EP AO NC
      1      13.467        10.634      1      EP DN AQ NC
      1      17.631        14.759      1      EP PR
*******
      1                                1      AQ NC EP NC
      1                                1      AQ NC EP NP
      1                                1      AQ NC EP IN NC
      1                                1      AQ NC CC DN NC
      1                                1      AT AQ NC
      1                                1      CC DN NC
      1                                1      CC PE AQ
      1                                1      DN AC NC
      1                                1      DN AQ NC EP NP
      1                                1      DN NC EP PR
      1                                1      DN NC EP PD
      1                                1      DN NC EP AO NC
      1                                1      DN NC CC AO NC
      1                                1      DN NC EP AQ NC
      1                                1      DN PO KO
      1                                1      DN PO EP NC
      1                                1      DN PO EP NP
      1                                1      EP AQ AQ NC
      1                                1      EP AQ NC AQ
      1                                1      IN AQ NC AQ
      1                                1      IN AQ AQ NC
      1                                1      IN AQ AQ AQ NC
```

```
       1                          1      IN AQ AQ NC AQ
       1                          1      IN AQ NC EP DN NC
       1                          1      IN AQ NC EP DN AQ NC
       1                          1      IN NC EP PR
       1                          1      IN NC EP AD NC
       1                          1      IN NC EP IN NC
       1                          1      IN NC EP DN NC
       1                          1      IN NC EP IN NC AQ
       1                          1      IN NC EP IN AQ NC
       1                          1      IN NC EP DN AQ NC
       1                          1      IN NC AQ CC IN NC AQ
       1                          1      IN PI EP NC
       1                          1      NC AQ AQ
       1                          1      NC AQ EP NP
       1                          1      NC CC NC
**GROUP  37  5.919    158.015
*******
       1                          1      NC EP PR
       1                          1      NP CC NC
       1                          1      NP CC PE
       1                          1      NP EP NP
       1                          1      PE CC PE
       1                          1      PE CC DN NC
       1                          1      PE LC
       1                          1      PI EP IN NC
**GROUP   8  2.002     15.095
          1133.2177 1133.2177            Residual

   5264   5264      7087.5209

Degrees of freedom =  35

Chi-square/degrees of freedom =  202.5006
Groups =  9
```

Probabilistic lexical disambiguation. Grammar I. In Section III, we mentioned that the ambiguity created by multiple classifications of words in the dictionary (called lexical ambiguity) was, for the most part, resolved by Grammar I. However, 18 types, representing 50 tokens (19 types before consolidation) remained that were lexically ambiguous and not resolved by Grammar I.

As in Smith (1972), we found that a plausible method of removing lexical ambiguity from a corpus involved the use of a probabilistic grammar. The method consisted of these steps:

1. Given an ambiguous corpus and a grammar, obtain estimates for the parameters associated with the rules of the grammar, as described above. If the lexical ambiguity remaining after resolution is small (in this case, it was less than 1 percent of the tokens), then the parameters will not be too inaccurate.

2. For each lexically ambiguous type, select the lexical reading that has the highest probability according to the parameters obtained in (1).

In Table 13 we give the 18 ambiguous types and include the resolution suggested by the above method. An examination of the actual noun phrases suggests that this method is reasonably accurate in removing spurious ambiguities.

TABLE 13
Lexical Disambiguation of Noun Phrases in Grammar I

Freq.	Ambiguous type	Resolved type	Prob.
7	DN,PE AQ,NC	DN NC	.06
7	DN,PE AQ,PI,AI	DN PI	.005
6	NC,PI	NC	.24
5	AQ,PI,AI,DV PD	AQ PD	.002
5	DN NC AQ,LC	DN NC AQ	.001
4	DN,PE PI,AI,LC	DN PI	.005
3	AQ,PI,AI, NC	AQ NC	.005
2	AQ,NC,KO NC,LC	AQ NC	.005
2	EP AQ,PI,AI NC	EP AQ NC	.0003
1	AD,PE AQ AQ,NC	AD AQ NC	.0003
1	AQ,LC AQ NC	AQ AQ NC	.0001
1	DN NC EP AQ,PI,AI,NC	DN NC EP AQ NC	.000009
1	DN,PE NC,LC	DN NC	.06
1	DN,PE NC,EP,LC	DN NC	.06
1	DN,PE AQ AQ,NC	DN AQ NC	.006
1	NC EP DN,PE AQ,NC	NC EP DN NC	.0005
1	PE AQ,VT,VA,LC	PE AQ	.002
1	PE,PI,AO	PE	.16

Grammar I and the sections of the noun-phrase corpus. Our noun-phrase corpus was drawn from three parts of the 33 sessions (Section I is Sessions 1 through 3, II is 14 through 16, and III is 31 through 33). We analyzed each of these sessions separately, especially since we were interested in the developmental aspects of the child and his growing sophistication in the use of language. Table 14 shows the results of analyzing each of the three different time sections of the three sessions separately. (The slight discrepency between Table 14 and Table 8 is due to a correction we made in the text that did not affect the fit.) For example, the first time section, Sessions 1-3, contained a total of 136 types in the corpus, with 104 types recognized. Eighty-seven of these types were ambiguous in the corpus, meaning that at least one word in the noun phrase had more than one grammatical classification. Of these 87, 73 were recognized and of the 73, 64 had the ambiguity resolved to a single lexical form. Of the remaining 9 types, 8 had the ambiguity reduced in a number of lexical forms. Similar remarks are to be made about the tokens, which are tabulated in the second column. The ambiguity factors for the three sections are also shown. What is important is the reduced ambiguity factor as defined above, which reflects the small amount of ambiguity in the final analysis. In other words, the reduced ambiguity factor of 21 indicates the number of tokens among the 1,562 recognized that have more than one derivation tree. For the second time period, Sessions 14-16, the reduced ambiguity factor is 9, and for the third time section, Sessions 31-33, the reduced ambiguity factor is 22.

After making the analysis shown in the table, we consolidated the types for each time section, and the subsequent analysis is in terms of these types. The first time section consolidated to 72 types, the second to 67, and the third to 89.

A small amount of lexical ambiguity remained in each section after resolution, and as in the corpus as a whole, we used the probabilistic method of disambiguation described above and consolidated the types after this disambiguation. After this reduction there were 64 types in the first section, 63 in the second, and 81 in the third.

There is a fairly substantial reduction in the percentage of types recognized in the third time portion. The first two are just about the same. The reduction in number of tokens recognized in the third time section is small, amounting to only a small percentage. The explanation for the smaller percentage of types recognized in Section 3 is that Section 3 is substantially larger than the first two sections (2351 tokens compared with 1629, 1566 tokens in Sections 1 and 2, respectively). We have noticed in the previous work that the larger the sample, the smaller the percentage of recognized types.

TABLE 14
Section Comparison of Noun-Phrase Corpus

	Section 1		Section 2		Section 3	
Description	Types	Tokens	Types	Tokens	Types	Tokens
Total	136	1629	126	1566	221	2351
Recognized	104	1562	98	1513	141	2191
% Recognized	76.47	95.8	77.78	96.62	63.80	93.19
Ambiguous	87	673	84	436	158	721
% Ambiguous	63.97	41.31	66.67	27.84	71.49	30.67
Amb. recog.	73	655	67	412	110	649
% Amb. recog.	83.91	97.33	79.76	94.50	69.62	90.01
Amb. resolved	64	636	61	403	98	627
Amb. reduced	8	18	5	6	11	19
Original ambiguity factor	1006		612		1188	
Parsed ambiguity factor	956		550		974	
Reduced ambiguity factor	21		9		22	
Types after consolidation	72		67		89	
Types after probabilistic disambiguation	64		63		81	

Some of the differences among the three sections are startling. Table 15 shows the most frequent types from each section, using the geometric model, together with the observed, expected, and chi-square contribution of each type. It is rather surprising to see DN NC occurred 394 times in Section I, and only 164 and 235 times in Sections 2 and 3, respectively. In Section I, EP NC occurred 93 times and in Section 3 it occurred 91 times.

TABLE 15
Section Comparison
High-frequency Noun Phrases

Observed	Expected	Chisquare	Count	Noun phrase type
Section I (Sessions 1-3)				
394	119.707	626.214	1	DN NC
206	199.028	.210	1	PE
191	52.473	363.071	1	PD
177	102.194	54.028	1	NP
97	387.618	217.143	1	NC
82	24.987	127.811	1	IN NC
60	63.357	.129	1	EP NC
39	57.122	5.436	1	AQ NC
38	26.943	4.136	1	PI
35	19.566	11.398	1	EP DN NC
33	13.834	25.184	1	PR
22	5.216	50.840	1	AO NC
22	17.641	.844	1	DN AQ NC
Section II (Sessions 14-16)				
427	227.669	173.645	1	PE
302	127.523	237.352	1	PD
164	60.561	174.971	1	DN NC
93	61.828	15.216	1	EP NC
85	71.429	2.392	1	NP
85	40.009	49.474	1	PR
62	28.300	38.947	1	IN NC
47	22.410	25.895	1	PI
39	13.225	48.306	1	AO NC

34	406.461	340.389	1	NC
21	6.180	33.182	1	AD NC
16	7.916	7.267	1	IN AQ NC
15	16.939	.122	1	DN AQ NC
Section III (Sessions 31-33)				
795	333.957	635.110	1	PE
337	179.686	136.853	1	PD
235	105.328	158.415	1	DN NC
137	44.373	191.271	1	IN NC
119	96.680	4.924	1	PR
91	82.134	.852	1	EP NC
73	52.019	8.064	1	PI
62	27.989	40.123	1	NP
46	18.694	38.438	1	AO NC
43	620.679	536.728	1	NC
30	12.385	23.653	1	IN AQ NC
18	29.397	4.039	1	DN AQ NC
14	7.876	4.017	1	AD NC

As an interesting comparison, Suppes (1970, 1971) in previous work on ADAM 1, found that in the earliest stages of development of an English-speaking child, the noun by itself is the most frequently occurring noun phrase. Also, in the study of the ERICA corpus (Smith 1972), the preponderance on the noun occurring alone decreases as the child nears three years of age. Quite plausibly, for a French-speaking child, the noun does not occur by itself (even at first), but rather with an article, which is natural in French for obvious reasons. Therefore, we see in Philippe's speech the same tendency that we have seen before in English corpora, but with the straightforward addition of the required article.

More general remarks of a developmental nature are reserved until the end of the section on Grammar II.

V. DICTIONARY II

Our second analysis, built directly on the first analysis, is aimed at an approach less based on the earlier work with English. We especially use the ideas found in Dubois (1970) to describe the construction of the determiner phrase that modifies a noun. These rules of the grammar are examined in more detail in Section VI.

To accommodate Dubois' ideas, Dictionary I required some modifications. In passing to Dictionary II, we have changed the grammatical categories as little as possible, in order to maintain a meaningful comparison. This means that some of our abbreviations, etc., still make Dictionary II look very much like Dictionary I. We have, however, at critical points introduced changes.

A full set of grammatical categories comparing Dictionary I with Dictionary II is shown in Table 16.

TABLE 16
Correspondence between Dictionary I and II.

Dictionary I	Dictionary II
IN	IN
DN	DN
PE	PE
PD	PD
PR	PR
PI1-7	PI1-7

PI8	DT QA NU
PO	PO
AQ	AQ
AC	AC
AN	AQ
AD	AD
AT	AT
AO	AO
AI	DT QA NU
NC	NC
NP	NP
DV	DV QR QA
EP	EP
CS	CS
CC	CC
KO	KO
LC	LC
MS	MS

First, to facilitate the rewrite rules of Grammar II described in the next section, we broke the category of PI8 of Dictionary I into three categories, *denotative* (DN), the category that Dubois calls *numerical* (NU) and Dubois' category of *absolute quantitative* (QA). Associated with this last category is also Dubois' category of *relative quantitative* (QR), which is a subdivision of DV in Dictionary I. Note that in the categories QA and QR we classified words which were classified as indefinite pronouns or possibly as adjectives or adverbs in Dictionary I. In one point we did not follow Dubois completely, for we still entered many of the words he classifies as QR also as adverbs in Dictionary II. For example, beaucoup was coded both as QR and DVO. Also, to anticipate Grammar II, in the rewrite rules using QR the category of lexical items in the category QR are ordinarily followed by de.

As might be expected, related to the recoding in Dictionary II of the category pronom inde'fini (PIS), there is a corresponding recoding from Dictionary I to Dictionary II of the category of adjectifs inde'finis (AI) into the same three categories DT, AQ, and NU.

The third major change to be noted in passing from Dictionary I to Dictionary II is the classification of adverbs (DV). In Dictionary II this classification is broken up into the adverb category, the absolute quantitative category and the relative quantitative category. We have already remarked about the two latter categories in Dictionary II, and consequently, nothing more is said here.

However, because the number of lexical items in these new categories is small, and because there is a lack of systemization of all the words to be considered by Dubois, we list the complete category for DT, NU, QA and QR. The following four words were placed in the denotative category DT: *autre(s)*, *la' (when following a noun), me'me(s)*. The numerical category NU contained the words: *chaque*, *nulle*, *quelque*, *plusieurs*. The absolute quantitative category contained the following words: *tout*,*toute(s)*, *tous*. Finally, the relative quantitative category contained the following words: *assez*, *beaucoup*, *peu*,*plus*,*trop*.

The coding of *la'* was DT when it followed a noun and seemed to have the interpretation of the composed form of the demonstrative adjective in such sentences as *Appuie sur le bouton la'*. The ordinal numbers have been coded as qualificative adjectives, that is, category AQ, following Dubois (1970, p. 54).

Some of the decisions made in establishing the categories of Dictionary II will be more evident when we turn to the production rules of Grammar II.

VI. NOUN-PHRASE GRAMMAR II

Grammar II, which is derived from Dubois (1970), is closer than our Grammar I to the current literature on generative grammars of the French language. It is already clear from discussing the construction of Dictionary II how the grammatical categories have been changed. In order to facilitate the comparison of the two grammars, the main subheadings of this section are precisely the same as those of the section dealing with Grammar I.

Nonterminal vocabulary of Grammar II. In addition to the grammatical categories introduced in Dictionary II, we have used the following additional nonterminal symbols in Grammar II. In the case of Grammar II, these additional symbols cannot be rewritten by lexical rules, but they occur in derivation trees always at least two nodes from a terminal node. As in Grammar I, the label of the root of all the noun phrases is SN (*syntagme nominal*). Following Dubois we also introduced GN and, for more specialized purposes of the final probabilistic analysis, GN'. The nonterminal of GN is an abbreviation for *groupe nominal*. We also have NP* playing a similar role to NP* in Grammar I.

Additional nonterminals taken from Dubois are PREART for pre-articles, POSTART for post-articles, ART for articles, and NUM for numerals. As in the case of Grammar

I, we also used ADJP for adjective phrases in pre-position and POSTADJP for adjective phrases in post-position. Finally, we used N for nouns with N being rewritten as either a common noun or a proper noun, that is, as a NC or NP.

As in Grammar I, by again omitting the subscripts introduced in the original dictionary coding, we achieved the same level of abstraction for the two grammars.

The production rules. The 56 production rules of Grammar II are shown in Table 17. As in the case of rules of Grammar I, the production rules have a surprisingly standard character. The prototypes of a good many of them are to be found in Dubois (1970).

In our formalization of the notions of Dubois, we first used a grammar somewhat closer to that of Dubois. We discuss below the reasons for the main modifications to the grammar implicit in Dubois.

TABLE 17
Production Rules of Grammar II

Label	Rule
(1,1)	SN -> GN'
(1,2)	SN -> GN EP GN
(1,3)	SN -> GN CC GN
(1,4)	SN -> NC
(1,5)	SN -> CC GN
(1,6)	SN -> KO GN

(1,7) SN -> GN KO
(1,8) SN -> LC GN
(1,9) SN -> GN LC
(1,10) SN -> EP NC
(1,11) SN -> EP GN'

(2,1) GN -> GN'
(2,2) GN -> NC

(3,1) GN' -> PE
(3,2) GN' -> PD
(3,3) GN' -> PR
(3,4) GN' -> PI
(3,5) GN' -> PO
(3,6) GN' -> NP*
(3,7) GN' -> DN PO
(3,8) GN' -> D NP*
(3,9) GN' -> D
(3,10) GN' -> NP
(3,11) GN' -> D NC
(3,12) GN' -> D NP

(4,1) D -> PREART AD POSTART
(4,2) D -> PREART ART POSTART
(4,3) D -> AD POSTART
(4,4) D -> POSTART
(4,5) D -> PREART AD
(4,6) D -> PREART POSTART
(4,7) D -> ART POSTART
(4,8) D -> PREART ART
(4,9) D -> AD
(4,10) D -> ART
(4,11) D -> PREART

(5,1) ART -> AO
(5,2) ART -> AT
(5,3) ART -> IN
(5,4) ART -> DN

(6,1) POSTART -> NUM
(6,2) POSTART -> NUM DT
(6,3) POSTART -> DT

(7,1) NUM -> AC
(7,2) NUM -> NU

(8,1) PREART -> QA
(8,2) PREART -> QR EP

(9,1) NP* -> ADJP N POSTADJP
(9,2) NP* -> ADJP N
(9,3) NP* -> N POSTADJP

(10,1) ADJP -> AQ
(10,2) ADJP -> ADJP AQ

(11,1) POSTADJP -> AQ
(11,2) POSTADJP -> POSTADJP AQ

(12,1) N -> NC
(12,2) N -> NP

The basic structure of the production rules reflects Dubois' structural representation of the nominal group (GN) as the determiner plus a noun, where the determiner itself is rewritten as shown in rules (4,1) to (4,11). Thus, the structure of the determining phrase and its importance to the noun phrase is a critical part of the Dubois treatment. We have taken over Dubois' distinctions of pre-article, demonstrative articles and post-articles, as mentioned above. As the production rules for pre-articles we have also used his division into absolute and relative quantities as reflected in rules (8,1) and

(8,2). Unlike Dubois, we have also coded as adverbs the words coded as QR. We adapted his rewriting rules for the post-articles, and this is reflected in rules (6,1) (6,2) and (6,3). Dubois included as one of the possible rewrite rules for post-articles the introduction of the word tel, but we omitted this rule because tel does not appear in the Philippe corpus.

Examination of Table 17 shows that we have a larger number of groups of rules in Grammar II than in Grammar I; in particular there are 12 groups of rules in Grammar II compared to 7 in Grammar I. In the first group we have 11 rules compared to 9 in the first group of Grammar I. These are the rules that govern the first step from the labeled root of the tree SN.

The two rules in the second group, starting from GN, generate either symbol GN' or the single terminal NC. The reason for introducing these rules was to generate the terminal NC (common nouns) at a high level in the grammar in order to obtain a better probabilistic fit. Comparing this feature of our Grammar II with our original "Dubois" grammar, we see that most of our modifications have mainly changed the level of introduction of certain symbols, and not the general nature of the grammar or to any degree the noun phrases recognized.

Rule (1,10) requires separate remarks. The purpose of this rule is to generate, at a high level in the grammar, phrases like d'argent. In these phrases, the preposition EP is in fact the partitive construction, as is indicated by the common English translation some money. Since our original "Dubois" grammar does not allow for this construction, we added the following rule to handle these cases: Our first efforts to add it did so by adding as a rule to group 3

GN' -> EP NP* .

The problem with this rule, from a probabilistic point of view, is that it does not allow for the rather common occurrence of EP NC (which occurred 244 times in the consolidated version of Philippe's speech under Grammar II), as opposed to phrases like EP NC AQ, which occurred only 6 times. Therefore, we changed the grammar to allow for the separate generation of EP NC, and took care to see that no ambiguities were introduced.

The third group of rules, (3,1) to (3,12), provide the rewrite rules for the nominal group corresponding fairly closely to suggestions in Dubois. There are however several differences. Dubois (1970, p. 35) discusses the view that the GN structure has these forms:

GN -> D N
GN -> NP
GN -> PRONOUN

(The general category PRONOUN is Dubois', not ours.) The first change that we made was to add rules (3,1) to (3,5) to accommodate the several kinds of pronouns that we have offered. This change is hardly more than terminology. A deeper change, but one that is consistent with Dubois' treatment of adjectives (see pp. 126-131), is to allow adjective phrases to be introduced after the determiner element, using rules (9,1) through (9,3). Thus, we have added rules (3,6) and (3,8) to the rules in group 3.

One additional rule that deserves mention is rule (3,7), which is also a departure from Dubois. This rule was introduced to remedy a rather bad probabilistic discrepancy that had occurred in Grammar I. Noun phrases of the type DN PO can be generated in Grammar I by the use of rules (3,15), (2,7), and (6,4). However, as Table 10 shows, the theoretical frequency of this type of utterance was negligible, while the type has an actual frequency of 12 in our selected nine hours of the corpus. The much better fit of the new rule introduced in Grammar II is seen from the data in Table 20 below. The predicted frequency is now 13.805, which matches closely the observed frequency of 12.

The fourth group of rules, the rewrite rules for determiners, constitute the most important contribution of Dubois' analysis to Grammar II. Here we have followed closely his suggestions, because we think they provide an excellent framework for handling the various parts of speech that precede nouns in noun phrases. We can summarize these rules by saying that they generate the 11 grammatically sensible possibilities out of the 16 possibilities obtained by either using or not using each of the categories PREART, AD, ART, POSTART.

The fifth group of rules are the rewrite rules for articles. The sixth group of rules are for the post-articles to be rewritten as "numerical" terms, which can lead to a single word in the NUM category or NUM followed by a denotative or a denotative alone. The seventh group of rules are the rewrite rules for splitting the numerical category NUM into the grammatical categories AC and NU, which are discussed above in the section on Dictionary II. The pre-article rewrite rules form the eighth group and provide for rewriting pre-articles as either absolute or relative quantities, with of course the relative quantities followed by a preposition. The ninth, tenth, eleventh and twelfth groups of rules are close to rules occurring in Grammar I and do not require additional comment.

Nonprobabilistic fit of Grammar II. The data are summarized in Table 18. Grammar II does not do quite as well as Grammar I in terms of the total number of noun phrases recognized. Grammar II parses 5,145 tokens in contrast to the 5,266 for Grammar I. The number of types that are parsed is also smaller -- 212 for Grammar II, 223 for Grammar I. The reduced ambiguity factor is also four times larger for Grammar II than for Grammar I, namely, 241 for Grammar II compared to 52 for Grammar I.

TABLE 18
Nonprobabilistic Fit of Grammar II Noun-phrase Corpus

Description	Types	Tokens
Total	347	5547
Recognized by Grammar II	212	5145
Percent recognized	61.1%	92.8%
Lexically ambiguous	236	1804
Percent lexically ambiguous	68.0%	32.5%
Ambiguous recognized	149	1634
Percent ambiguous recognized	63.1%	90.6%
Ambiguity resolved	133	1401
Ambiguity reduced	10	21
Original ambiguity factor	2607	
Parsed ambiguity factor	2203	
Reduced ambiguity factor	241	

Since the entries of Table 18 correspond exactly to the entries of Table 8 in our discussion of Grammar I in Section 4, so we will not repeat the explanation. Notice that the number of types originally present in the corpus is somewhat different here (347 compared to 345). This of course is because the change to Dictionary II produces a change in the categorization of noun phrases.

After initial parsing, the 212 recognized types consolidated to 140 types. This compares to a somewhat higher figure of 143 types for Grammar I.

It might be thought from this discussion that Grammar II is not as satisfactory as Grammar I. However, under the much tighter and stricter probabilistic criterion that we next discuss Grammar II generates a language that has a much better fit to the corpus of noun phrases than does that generated by Grammar I.

Probabilistic fit of Grammar II. Table 19 gives the parameters for the standard fit of Grammar II, obtained by the maximum-likelihood method, and Table 20 shows the results of the goodness-of-fit test for each noun-phrase type. Once again, types with expected frequency less than 5 were grouped together to avoid the problem of large chi-square contributions, and we used a continuity correction of .5.

TABLE 19

Parameters for Grammar II

Philippe's Noun Phrases*

Label	Prob.	Rule
(1,1)	.8603	SN -> GN'
(1,2)	.0249	SN -> GN EP GN
(1,3)	.0056	SN -> GN CC GN
(1,4)	.0338	SN -> NC
(1,5)	.0031	SN -> CC GN
(1,6)	.0000	SN -> KO GN
(1,7)	.0047	SN -> GN KO
(1,8)	.0000	SN -> LC GN
(1,9)	.0004	SN -> GN LC
(1,10)	.0474	SN -> EP NC
(1,11)	.0198	SN -> EP GN'
(2,1)	.7893	GN -> GN'
(2,2)	.2107	GN -> NC
(3,1)	.2996	GN' -> PE
(3,2)	.1792	GN' -> PD
(3,3)	.0501	GN' -> PR
(3,4)	.0316	GN' -> PI
(3,5)	.0000	GN' -> PO
(3,6)	.0162	GN' -> NP*
(3,7)	.0031	GN' -> DN PO
(3,8)	.0343	GN' -> D NP*
(3,9)	.0106	GN' -> D
(3,10)	.0807	GN' -> NP
(3,11)	.2930	GN' -> D NC
(3,12)	.0015	GN' -> D NP
(4,1)	.0000	D -> PREART AD POSTART
(4,2)	.0000	D -> PREART ART POSTART
(4,3)	.0000	D -> AD POSTART
(4,4)	.0313	D -> POSTART
(4,5)	.0000	D -> PREART AD
(4,6)	.0000	D -> PREART POSTART
(4,7)	.0331	D -> ART POSTART
(4,8)	.0074	D -> PREART ART
(4,9)	.0349	D -> AD
(4,10)	.8854	D -> ART
(4,11)	.0080	D -> PREART

(5,2)	.0086	ART -> AT
(5,3)	.2958	ART -> IN
(5,4)	.6956	ART -> DN
(6,1)	.4667	POSTART -> NUM
(6,2)	.0000	POSTART -> NUM DT
(6,3)	.5333	POSTART -> DT
(7,1)	.5918	NUM -> AC
(7,2)	.4082	NUM -> NU
(8,1)	.7600	PREART -> QA
(8,2)	.2400	PREART -> QR EP
(9,1)	.0206	NP* -> ADJP N POSTADJP
(9,2)	.7572	NP* -> ADJP N
(9,3)	.2222	NP* -> N POSTADJP
(10,1)	.9742	ADJP -> AQ
(10,2)	.0258	ADJP -> ADJP AQ
(11,1)	.9833	POSTADJP -> AQ
(11,2)	.0167	POSTADJP -> POSTADJP AQ
(12,1)	.9918	N -> NC
(12,2)	.0082	N -> NP

*Correction for continuity = .5; expected < 5 are grouped together.

TABLE 20

Noun-Phrase Types for Grammar II

Observed	Expected	Chisquare	Count	Noun-phrase type
1428	1326.235	7.732	1	PE
830	793.348	1.647	1	PD
793	798.645	.033	1	DN NC
324	357.099	2.976	1	NP
281	339.671	9.962	1	IN NC
244	244.000	.001	1	EP NC
237	221.806	.973	1	PR
174	174.000	.001	1	NC

```
    152        139.894         .963        1     PI
     60         18.405       91.755        1     EP DN NC
     52         29.101       17.240        1     IN AQ NC
     43         52.521        1.549        1     AQ NC
     40         68.424       11.396        1     DN AQ NC
     37         45.292        1.341        1     AD NC
     18         11.192        3.555        1     AC NC
     16         20.269         .701        1     DN NC AQ
     15         15.918         .011        1     DN DT NC
*******
     23                                    1     IN NC EP NC
     21                                    1     DN NC EP NP
     13                                    1     DN NC EP NC
**GROUP  57      6.635      374.734
     13         12.295         .003        1     IN
     13          8.620        1.746        1     IN NC AQ
     12         13.805         .123        1     DN PO
     11          8.230         .626        1     EP NP
     11          7.719        1.002        1     NU NC
*******
     12                                    1     DN DT
     11                                    1     CC PD
     10                                    1     DN NC KO
**GROUP  33      6.258      110.037
     10          5.681        2.567        1     NC EP NC
      9          6.770         .442        1     IN DT NC
      9          5.041        2.374        1     QA DN NC
*******
     10                                    1     IN DT
      9                                    1     EP AQ NC
      8                                    1     AD AQ NC
**GROUP  27      5.336       83.946
      8          9.879         .192        1     AT NC
      8         15.558        3.202        1     NC AQ
*******
      8                                    1     NP CC NP
      8                                    1     PD KO
      7                                    1     DN NC EP DN NC
**GROUP  23      6.110       43.968
*******
      6                                    1     EP NC AQ
      6                                    1     QA
      5                                    1     DN NP
      5                                    1     IN AQ NC EP NC
      5                                    1     QR EP NC
**GROUP  27      7.230       51.366
*******
      4                                    1     AD NC KO
      4                                    1     CC PE
      4                                    1     DN NC CC DN NC
      4                                   1      IN NC CC IN NC
      3                                   1      AC AQ NC
**GROUP  19      5.632       29.403
*******
      3                                   1      DN NC EP PE
      3                                   1      DN NC EP AD NC
      3                                   1      DN NC EP IN NC
**GROUP   9      5.564        1.550
*******
      3                                   1      EP AT NC
      3                                   1      EP NU NC
      3                                   1      IN EP NU NC
      3                                   1      NC EP NP
      3                                   1      NU AQ NC
      3                                   1      PD CC PD
      3                                   1      QA DN NC EP NP
      2                                   1      AD NP
      2                                   1      AQ AQ NC
      2                                   1      DN AC AQ NC
**GROUP  27      5.657       76.789
      2         21.613       16.902       1      DT NC
      2          7.828        3.626       1      EP IN NC
*******
      2                                   1      DN AQ NC AQ
      2                                   1      DN NC CC DN NC AQ
      2                                   1      EP DT
      2                                   1      EP DT NC
      2                                   1      IN AC NC
**GROUP  10      5.865        2.252
*******
      2                                   1      IN EP NC
      2                                   1      NC EP DN NC
      2                                   1      NP AQ
      2                                   1      PD EP NP
**GROUP   8      5.182        1.037
*******
      2                                   1      PD EP PD
      2                                   1      PE LC
      1                                   1      AC NC AQ
      1                                   1      AC NC CC AC NC
      1                                   1      AD AQ NC KO
      1                                   1      AD NC AQ
      1                                   1      AQ NC EP NC
      1                                   1      AQ NC EP NP
      1                                   1      AQ NC EP IN NC
      1                                   1      AQ NC CC DN NC
      1                                   1      AT
**GROUP  13      5.283        9.857
      1          8.243        5.516       1      DN AC NC
*******
```

```
      1                                    1      AT AQ NC
      1                                    1      CC DN NC
      1                                    1      DN AQ NC EP NP
      1                                    1      DN NC EP PR
      1                                    1      DN NC EP PD
**GROUP    5     6.525        .161
      1          5.112       2.552         1      EP PR
*******
      1                                    1      DN NC EP DT NC
      1                                    1      DN PO KO
      1                                    1      DN PO EP NC
      1                                    1      DN PO EP NP
      1                                    1      EP AQ AQ NC
      1                                    1      EP AQ NC AQ
      1                                    1      EP DN AQ NC
      1                                    1      IN AQ NC AQ
      1                                    1      IN AQ AQ AQ NC
      1                                    1      IN AQ NC EP DN NC
      1                                    1      IN AQ NC EP DN AQ NC
      1                                    1      IN DT AQ NC
      1                                    1      IN DT EP NC
      1                                    1      IN DT AQ NC AQ
      1                                    1      IN NC EP PR
      1                                    1      IN NC EP AD NC
      1                                    1      IN NC EP IN NC
      1                                    1      IN NC EP DN NC
**GROUP   18     5.296      28.123
*******
      1                                    1      IN NC EP IN NC AQ
      1                                    1      IN NC EP IN AQ NC
      1                                    1      IN NC EP DN AQ NC
      1                                    1      IN NC AQ CC IN NC AQ
      1                                    1      NC AQ AQ
      1                                    1      NC AQ EP NP
      1                                    1      NC CC NC
      1                                    1      NC EP PR
      1                                    1      NP CC NC
      1                                    1      NP CC PE
      1                                    1      NP EP NP
      1                                    1      PE CC PE
**GROUP   12     5.750       5.750
*******
      1                                    1      PE CC DN NC
      1                                    1      QA EP IN NC
      1                                    1      QR EP AQ NC
**GROUP    3     1.190       1.443

               269.2729    269.2729               Residual
```

5145	5145	1282.4002	Totals

Degrees of freedom = 11

Chi-square/degrees of freedom = 116.5818

Groups = 15

We shall compare the fit as shown in Table 20 with the fit of Grammar I as shown in Table 10 of Section IV, to substantiate for our claim that Grammar II has a better fit than does Grammar I. Notice that the residual of Grammar II is only 269, whereas that of Grammar I is 1112. This means that over 20 percent of the probability in Grammar I was distributed among noun-phrase types that did not occur in Philippe, while in Grammar II only about 5 percent of the probability is so distributed. This is not a measure of the adequacy of a grammar, but instead of the tightness of the grammar relative to the corpus under consideration. Grammar I, primarily our previous experience with English, is too broad to provide a sharp characterization of the speech of Philippe. The slight decrease in the number of types/tokens recognized by Grammar II over Grammar I is more than compensated for in this residual.

The goodness-of-fit of Grammar II is even more encouraging. With 11 degrees of freedom, the chi-square sum over the noun-phrase types is 1282 for Grammar II, compared to 5527 for Grammar I. On comparing the high frequency noun phrases for the two grammars we quickly see that the predictions of Grammar II are much more accurate than those of Grammar I. The high-frequency noun-phrase types (which happen to be the same in both dictionaries) are shown in Table 21.

TABLE 21

Comparison of Grammars I and II

Philippe Noun phrases

Observed	Chi-square contributions Grammar I	 Grammar II	Noun-phrase
1428	406.1	7.7	PE
830	225.6	1.6	PD
793	609.0	.0	DN NC
324	41.3	3.0	NP
281	151.2	10.0	IN NC
244	233.3	.0	EP NC
237	67.7	1.0	PR
174	960.0	.0	NC
156	24.9	1.0	PI

Total observed (in this sample) = 4467

There are other types not so well handled by Grammar II, but among the high-frequency types shown in Table 21, which account for 4467 noun phrases out of 5547 in the corpus, it is clear that Grammar II is the better fit.

We now turn to the geometric fit of Grammar II. Table 22 gives the parameters of that fit, and Table 23 gives the details of the fit for each noun-phrase type. (These tables correspond to Tables 11 and 12 of Section IV.)

TABLE 22
Parameters for Geometric Fit of Grammar II

(1,1)	4426	3810.5897	SN -> GN'
(1,10)	244	988.3168	SN -> EP NC
(1,4)	174	256.3305	SN -> NC
(1,2)	128	66.4820	SN -> GN EP GN
(1,11)	102	17.2428	SN -> EP GN'
(1,3)	25	4.4721	SN -> GN CC GN
(1,7)	24	1.1599	SN -> GN KO
(1,5)	16	.3008	SN -> CC GN
(1,9)	2	.1053	SN -> GN LC

Theta = .7406394

(2,1)	281	281.0000	GN -> GN'
(2,2)	75	75.0000	GN -> NC

Theta = .7893258

(3,1)	1441	1781.9908	GN' -> PE
(3,11)	1409	1121.6682	GN' -> D NC
(3,2)	862	706.0304	GN' -> PD
(3,10)	388	444.4085	GN' -> NP
(3,3)	241	279.7315	GN' -> PR
(3,8)	165	176.0761	GN' -> D NP*
(3,4)	152	110.8305	GN' -> PI
(3,6)	78	69.7619	GN' -> NP*
(3,9)	51	43.9114	GN' -> D
(3,7)	15	27.6399	GN' -> DN PO
(3,12)	7	46.9510	GN' -> D NP

```
Theta =        .3705533

(4,10)   1445     1289.6780       D -> ART
(4,9)      57      270.5178       D -> AD
(4,7)      54       56.7428       D -> ART POSTART
(4,4)      51       11.9021       D -> POSTART
(4,11)     13        2.4965       D -> PREART
(4,8)      12         .6627       D -> PREART ART

Theta =        .7902439

(5,4)    1051     1148.3907       ART -> DN
(5,3)     447      275.5905       ART -> IN
(5,2)      13       87.0189       ART -> AT

Theta =        .7600203

(6,3)      56       56.0000       POSTART -> DT
(6,1)      49       49.0000       POSTART -> NUM

Theta =        .5333333

(7,1)      29       29.0000       NUM -> AC
(7,2)      20       20.0000       NUM -> NU

Theta =        .5918367

(8,1)      19       19.0000       PREART -> QA
(8,2)       6        6.0000       PREART -> QR EP

Theta =        .7600000

(9,2)     184      191.5033       NP* -> ADJP N
(9,3)      54       40.5835       NP* -> N POSTADJP
(9,1)       5       10.9132       NP* -> ADJP N POSTADJP

Theta =        .7880795

(10,1)    189      189.0000       ADJP -> AQ
(10,2)      5        5.0000       ADJP -> ADJP AQ

Theta =        .9742268

(11,1)     59       59.0000       POSTADJP -> AQ
(11,2)      1        1.0000       POSTADJP -> POSTADJP AQ

Theta =        .9833333

(12,1)    241      241.0000       N -> NC
(12,2)      2        2.0000       N -> NP
```

Theta = .9917695

TABLE 23
Noun Phrase Types for Grammar II
Geometric Fit

Observed	Expected	Chisquare	Count	Noun-phrase type
1428	1412.027	.17	1	PE
830	559.449	130.356	1	PD
793	533.812	125.362	1	DN NC
324	352.144	2.170	1	NP
281	128.104	181.294	1	IN NC
244	988.317	559.804	1	EP NC
237	221.656	.994	1	PR
174	256.330	26.123	1	NC
152	87.821	46.174	1	PI
52	15.312	85.523	1	IN AQ NC
43	42.092	.004	1	AQ NC
40	63.806	8.513	1	DN AQ NC
37	147.325	81.871	1	AD NC
*******				
60			1	P DN NC
23			1	N NC EP NC
21			1	N NC EP NP
18			1	C NC
**GROUP 122	5.114	2648.96		
16	13.648	.251	1	DN NC AQ
15	12.526	.311	1	DN DT NC
13	5.015	11.171	1	IN
*******				
13			1	N NC EP NC
13			1	N NC AQ
12			1	N DT
**GROUP 38	5.314	194.926		
12	21.901	4.036	1	DN PO
*******				
11			1	C PD
11			1	P NP
11			1	U NC
10			1	N NC KO
10			1	N DT

```
    10                                     1       C EP NC
**GROUP  63     6.060      525.697
      8        17.610        4.713          1       AD AQ NC
      8        40.449       25.236          1       AT NC
      8         9.003         .028          1       NC AQ
      5        22.344       12.698          1       DN NP
*******
      9                                     1       P AQ NC
      9                                     1       N DT NC
      9                                     1       A DN NC
      8                                     1       P CC NP
      8                                     1       D KO
      7                                     1       N NC EP DN NC
      6                                     1       P NC AQ
      6                                     1       A
      5                                     1       N AQ NC EP NC
      5                                     1       R EP NC
      4                                     1       D NC KO
      4                                     1       C PE
      4                                     1       N NC CC DN NC
**GROUP  84     5.006     1230.79
      2         6.167        2.180          1       AD NP
*******
      4                                     1       N NC CC IN NC
      3                                     1       C AQ NC
      3                                     1       N NC EP PE
      3                                     1       N NC EP AD NC
      3                                     1       N NC EP IN NC
      3                                     1       P AT NC
      3                                     1       P NU NC
      3                                     1       N EP NU NC
      3                                     1       C EP NP
      3                                     1       U AQ NC
      3                                     1       D CC PD
      3                                     1       A DN NC EP NP
      2                                     1       D AQ NC
**GROUP  39     5.290      208.508
*******
      2                                     1       N AC AQ NC
      2                                     1       N AQ NC AQ
      2                                     1       N NC CC DN NC AQ
      2                                     1       T NC
**GROUP   8     7.809         .012
*******
      2                                     1       P DT
      2                                     1       P DT NC
      2                                     1       P IN NC
      2                                     1       N AC NC
      2                                     1       N EP NC
      2                                     1       C EP DN NC
      2                                 1       P AQ
      2                                 1       D EP NP
      2                                 1       D EP PD
**GROUP  18     5.245       28.630
*******
      2                                 1       E LC
      1                                 1       C NC AQ
      1                                 1       C NC CC AC NC
      1                                 1       D AQ NC KO
      1                                 1       D NC AQ
      1                                 1       Q NC EP NC
      1                                 1       Q NC EP NP
      1                                 1       Q NC EP IN NC
      1                                 1       Q NC CC DN NC
      1                                 1       T
**GROUP  11     5.615        4.249
      1         6.487        3.834       1      DN AC NC
*******
      1                                 1       T AQ NC
      1                                 1       C DN NC
      1                                 1       N AQ NC EP NP
      1                                 1       N NC EP PR
**GROUP   4     5.270         .112
*******
      1                                 1       N NC EP PD
      1                                 1       N NC EP DT NC
      1                                 1       N PO KO
      1                                 1       N PO EP NC
      1                                 1       N PO EP NP
      1                                 1       P AQ AQ NC
      1                                 1       P AQ NC AQ
      1                                 1       P DN AQ NC
      1                                 1       P PR
      1                                 1       N AQ NC AQ
      1                                 1       N AQ AQ AQ NC
      1                                 1       N AQ NC EP DN NC
      1                                 1       N AQ NC EP DN AQ NC
      1                                 1       N DT AQ NC
      1                                 1       N DT EP NC
      1                                 1       N DT AQ NC AQ
      1                                 1       N NC EP PR
      1                                 1       N NC EP AD NC
      1                                 1       N NC EP IN NC
      1                                 1       N NC EP DN NC
      1                                 1       N NC EP IN NC AQ
      1                                 1       N NC EP IN AQ NC
      1                                 1       N NC EP DN AQ NC
      1                                 1       N NC AQ CC IN NC AQ
      1                                 1       C AQ AQ
      1                                 1       C AQ EP NP
```

```
     1                                                  1      C CC NC
     1                                                  1      C EP PR
     1                                                  1      P CC NC
**GROUP  29     5.006        110.275
*******
     1                                                  1      P CC PE
     1                                                  1      P EP NP
     1                                                  1      E CC PE
     1                                                  1      E CC DN NC
     1                                                  1      A EP IN NC
     1                                                  1      R EP AQ NC
**GROUP   6      1.015        19.807
               124.9099     124.9099                           Residual


 5145        5145          6409.6993

Degrees of freedom =           21

Chisquare/degrees of freedom =  305.2238
Groups =          11
```

When we use geometric distributions in each group of production rules, Grammar II again has a better fit than Grammar I in relation to the highest frequency noun-phrase types. More generally, Grammar II has, with geometric parameters, a chisquare sum of 6410 (21 degrees of freedom) compared to 7088 for Grammar I (35 degrees of freedom). The chisquare sum for Grammar II is in large part due to one group of noun-phrase types that contribute 2649 to the chisquare, so the fit is better than it looks. Moreover, the low residual of only 125 for Grammar II compared to 1133 of Grammar I indicates a tighter, more constrained grammar.

In all of the tables we have indicated not only the chi-square for the fit of the language generated by the grammar to the actual corpus, but we have also indicated the magnitude of chi-square divided by the number of net degrees of freedom. We have used this method as a rough and ready way of comparing the fit of different grammars with different numbers of degrees of freedom. For example, in the case of the geometric distribution we have 21 degrees of freedom rather than 11 in the case of the general Grammar II. However, when we look at the ratio of chi-square divided by degrees of freedom we see in spite of the greater degrees of freedom without even looking at a table of significance that the fit of the general Grammar II is better than the geometrically constrained grammar. We emphasize that we are not actually attaching a statistical significance for the purpose of emphasis to this ratio of chi-square to degrees of freedom. It is introduced for the purpose of easy, descriptive comparison.

Probabilistic lexical disambiguation, Grammar II. As in the case of Grammar I we used probabilistic methods to resolve the lexical ambiguity remaining in the corpus. There were 16 types (233 tokens) that were recognized with lexical ambiguity by Grammar II. In Table 24 we give the results of the probabilistic disambiguation. (Table 24 is similar to Table 13 in Section IV.)

TABLE 24
Lexcial Disambiguation of Noun Phrases in Grammar II

Freq.	Ambiguous type	Resolved type	Prob.
192	DN,PE	PE	.22
7	AC,NC	NC	.03
6	NC,PI	NC	.03
5	AD,PE	PE	.22
5	DN NC AQ,LC	DN NC AQ	.004
4	DN,PE DV,DT LC	DN DT	.0004
3	DN AC,NC	DN NC	.16
2	AQ,NC,KO NC,LC	AQ NC	.01
2	QA DN AC,NC	QA DN NC	.0008
1	AQ,LC AQ NC	AQ AQ NC	.0003
1	CC DN,PE	CC PE	.0006
1	DN,PE LC	PE LC	.00008
1	DN,PE NC,LC	DN NC	.16
1	DN,PE NC,EP,LC	DN NC	.16
1	NU,AQ NC	AQ NC	.01
1	PE,PO,AO	PE	.22

Grammar II and the sections of the noun-phrase corpus. Table 25 shows the comparison of the noun-phrases in the three sections of the noun-phrases that were used in this study. (Section I is Sessions 1 through 3, II is 14 through 16, and III is 31 through 33). We have applied the push-down store model with a geometric parameter for each group of rules to each of the sections. Our original intention was to look at the changes in the geometric parameters over time as evidence of development. Because the corpus itself does not show striking differences of a systematic kind over the three sections, we have not found any striking developmental trends at the more abstract level of the geometric parameters.

TABLE 25

Section Comparison of Noun-Phrase Corpus

	Section 1		Section 2		Section 3	
Description	Types	Tokens	Types	Tokens	Types	Tokens
Total	136	1629	126	1566	220	2351
Recognized	103	1528	90	1475	133	2142
% Recognized	75.74	93.80	71.43	94.19	60.45	91.11
Ambiguous	84	667	81	429	150	710
% Ambiguous	61.76	40.95	64.29	27.39	68.18	30.20
Amb. recog.	67	629	55	392	97	615
% Amb. recog.	79.76	94.30	67.90	91.38	64.67	86.62
Amb. resolved	60	597	50	330	86	476
Amb. reduced	5	9	2	2	6	10
Original ambiguity factor	938		573		1118	
Parsed ambiguity factor	881		498		846	
Reduced ambiguity factor	38		62		141	
Types after consolidation	75		67		89	
Types after probabilistic disambiguation	69		62		78	

The parameters and their predictions for the frequency of use of each rule in each section are shown in Table 26. The one thing to note is that the changes in the parameters across the three sections are not monotonic. We might, for example, expect for the more complex groups of rules that the geometric parameter would decrease in size because of the variety of rules used of a given type would increase with time, or put more exactly, the frequency of use would be more diverse. This is not the case for a number of groups of rules, as may be seen from the perusal of Table 26.

TABLE 26

Section Comparison of Geometric Parameters

Label	Observed	Expected	Rule
		Section I (Sessions 1-3)	
(1,1)	1249	1055.8730	SN -> GN'
(1,4)	97	326.2475	SN -> NC
(1,10)	60	100.8052	SN -> EP NC
(1,2)	51	31.1471	SN -> GN EP GN
(1,11)	49	9.6240	SN -> EP GN'
(1,3)	17	2.9736	SN -> GN CC GN
(1,7)	5	1.3296	SN -> GN KO
Theta =	.6910163		
(2,1)	109	109.0000	GN -> GN'
(2,2)	32	32.0000	GN -> NC

```
Theta =        .7730497

(3,11)     576       521.7718        GN' -> D NC
(3,10)     221       328.2780        GN' -> NP
(3,1)      211       206.5394        GN' -> PE
(3,2)      201       129.9463        GN' -> PD
(3,6)       55        81.7570        GN' -> NP*
(3,8)       45        51.4383        GN' -> D NP*
(3,4)       37        32.3629        GN' -> PI
(3,3)       34        20.3614        GN' -> PR
(3,9)       15        12.8106        GN' -> D
(3,7)        9         8.0599        GN' -> DN PO
(3,12)       3        13.6743        GN' -> D NP

Theta =        .3708399

(4,10)     598       564.3463        D -> ART
(4,7)       23        65.9320        D -> ART POSTART
(4,4)        5         7.7028        D -> POSTART
(4,8)        5          .8999        D -> PREART ART
(4,9)        4          .1051        D -> AD
(4,11)       4          .0139        D -> PREART

Theta =        .8831711

(5,4)      510       510.0000        ART -> DN
(5,3)      116       116.0000        ART -> IN

Theta =        .8146965

(6,3)       25        25.0000        POSTART -> DT
(6,1)        3         3.0000        POSTART -> NUM

Theta =        .8928571

(7,1)        3         3.0000        NUM -> AC

Theta =       1.0000000

(8,1)        6         6.0000        PREART -> QA
(8,2)        3         3.0000        PREART -> QR EP

Theta =        .6666667

(9,2)       79        81.8182        NP* -> ADJP N
(9,3)       20        14.8760        NP* -> N POSTADJP
(9,1)        1         3.3058        NP* -> ADJP N POSTADJP

Theta =        .8181818
```

```
(10,1)     80        80.0000      ADJP -> AQ
(10,2)      2         2.0000      ADJP -> ADJP AQ

Theta =       .9756098

(11,1)     21        21.0000      POSTADJP -> AQ
(11,2)      1         1.0000      POSTADJP -> POSTADJP AQ

Theta =       .9545455

(12,1)    100       100.0000      N -> NC

Theta =      1.0000000

                  Section II (Sessions 14-16)

(1,1)    1280      1117.1012      SN -> GN'
(1,10)     93       271.0571      SN -> EP NC
(1,4)      31        65.7702      SN -> NC
(1,11)     23        15.9587      SN -> EP GN'
(1,2)      20         3.8723      SN -> GN EP GN
(1,5)      10          .9396      SN -> CC GN
(1,7)      10          .2280      SN -> GN KO
(1,3)       8          .0730      SN -> GN CC GN

Theta =       .7573567

(2,1)      59        59.0000      GN -> GN'
(2,2)      17        17.0000      GN -> NC

Theta =       .7763158

(3,1)     427       513.0590      GN' -> PE
(3,11)    320       319.7921      GN' -> D NC
(3,2)     318       199.3279      GN' -> PD
(3,10)     97       124.2420      GN' -> NP
(3,3)      87        77.4406      GN' -> PR
(3,4)      50        48.2691      GN' -> PI
(3,8)      44        30.0864      GN' -> D NP*
(3,9)      10        18.7530      GN' -> D
(3,6)       8        11.6888      GN' -> NP*
(3,7)       1        19.3411      GN' -> DN PO

Theta =       .3766953

(4,10)    304       281.6356      D -> ART
(4,9)      34        69.5537      D -> AD
(4,4)      22        17.1772      D -> POSTART
(4,7)      12         4.2422      D -> ART POSTART
(4,11)      2         1.3912      D -> PREART
```

```
Theta =        .7530364

(5,4)      210       232.1327       ART -> DN
(5,3)      100        61.6087       ART -> IN
(5,2)        6        22.2586       ART -> AT

Theta =        .7345972

(6,1)       22        22.0000       POSTART -> NUM
(6,3)       12        12.0000       POSTART -> DT

Theta =        .6470588

(7,1)       15        15.0000       NUM -> AC
(7,2)        7         7.0000       NUM -> NU

Theta =        .6818182

(8,2)        2         2.0000       PREART -> QR EP

Theta =       1.0000000

(9,2)       41        42.0952       NP* -> ADJP N
(9,3)       10         8.0181       NP* -> N POSTADJP
(9,1)        1         1.8866       NP* -> ADJP N POSTADJP

Theta =        .8095238

(10,1)      42        42.0000       ADJP -> AQ

Theta =       1.0000000

(11,1)      11        11.0000       POSTADJP -> AQ

Theta =       1.0000000

(12,1)      50        50.0000       N -> NC
(12,2)       2         2.0000       N -> NP

Theta =        .9615385
```

Section III (Sessions 31-33)

```
(1,1)      903      1691.4686       SN -> GN'
(1,10)      91       355.7702       SN -> EP NC
(1,2)       57        74.8299       SN -> GN EP GN
(1,4)       40        15.7391       SN -> NC
(1,11)      30         3.3104       SN -> EP GN'
(1,7)        9          .6963       SN -> GN KO
```

```
(1,5)       6         .1465      SN -> CC GN
(1,3)       4         .0308      SN -> GN CC GN
(1,9)       2         .0082      SN -> GN LC

Theta =       .7896679

(2,1)     113      113.0000      GN -> GN'
(2,2)      26       26.0000      GN -> NC

Theta =       .8129496

(3,1)     803      824.5376      GN' -> PE
(3,11)    513      492.2491      GN' -> D NC
(3,2)     343      293.8728      GN' -> PD
(3,3)     120      175.4421      GN' -> PR
(3,8)      76      104.7390      GN' -> D NP*
(3,4)      71       62.5292      GN' -> PI
(3,10)     70       37.3299      GN' -> NP
(3,9)      26       22.2860      GN' -> D
(3,6)      15       13.3047      GN' -> NP*
(3,7)       5        7.9429      GN' -> DN PO
(3,12)      4       11.7666      GN' -> D NP

Theta =       .4029998

(4,10)    543      477.1134      D -> ART
(4,4)      24      109.3635      D -> POSTART
(4,7)      19       25.0682      D -> ART POSTART
(4,9)      19        5.7461      D -> AD
(4,8)       7        1.3171      D -> PREART ART
(4,11)      7         .3917      D -> PREART

Theta =       .7707809

(5,4)     331      396.2553      ART -> DN
(5,3)     231      120.3006      ART -> IN
(5,2)       7       52.4442      ART -> AT

Theta =       .6964065

(6,1)      24       24.0000      POSTART -> NUM
(6,3)      19       19.0000      POSTART -> DT

Theta =       .5581395

(7,2)      13       13.0000      NUM -> NU
(7,1)      11       11.0000      NUM -> AC

Theta =       .5416667
```

```
(8,1)      13      13.0000       PREART -> QA
(8,2)       1       1.0000       PREART -> QR EP

Theta =       .9285714

(9,2)      64      67.8644       NP* -> ADJP N
(9,3)      24      17.2537       NP* -> N POSTADJP
(9,1)       3       5.8819       NP* -> ADJP N POSTADJP

Theta =       .7457627

(10,1)     67      67.0000       ADJP -> AQ
(10,2)      3       3.0000       ADJP -> ADJP AQ

Theta =       .9571429

(11,1)     27      27.0000       POSTADJP -> AQ

Theta =      1.0000000

(12,1)     91      91.0000       N -> NC

Theta =      1.0000000
```

Given this absense of developmental trends in increasing complexity of noun phrases, and given the fact that the corpus spans what is an important developmental eriod in Philippe's speech, namely, the period running from 25 months to 39 months, it is natural to ask if there are any clear evidences of developmental trends in the corpus.

We conclude by exhibiting the statistic used extensively by Roger Brown and others, namely, the mean length of utterance. In Table 27 the mean length of utterance and the standard deviation, as well as the variance and the mode of the distribution, are shown for each hour of the 33 recording sessions. It is clear from this table that over the year and a half of the sessions there is a clear developmental trend in the mean length of utterances. As would be expected from sampling considerations, there is not a monotonic increase in the mean length of utterance across the 33 sessions, but the general trend is clear. This trend is toward increasing length of utterances. In order not to confuse this table with the many tables presented on noun phrases, it should be explicitly noted that the table for mean length of utterance deals with the entire length of utterance and not with the length of noun phrases. Given the lack of any developmental trends in the grammar of noun phrases, it is unlikely that a full utterance grammar will present any startling developmental trends, especially as we deal with the verb-phrase constructions. We plan to turn to these matters in the second report on the corpus of Philippe's speech.

TABLE 27

Utterance Length for 33 Sessions

Session	Mean	St.Dev.	Variance	Mode	No. of Utterances
1	3.12	1.59	2.52	2	418
2	3.09	1.77	3.13	2	488
3	3.49	1.92	3.6	2	493
4	3.24	1.70	2.8	3	639
5	2.91	1.76	3.0	1	311
6	3.16	2.39	5.7	3	472
7	3.62	1.83	3.3	3	357
8	3.36	1.98	3.9	2	206
9	3.24	1.78	3.1	3	421
10	3.05	1.88	3.53	1	415
11	3.76	2.45	6.00	1	479
12	3.52	2.26	5.13	1	463
13	4.07	2.34	5.46	1	416
14	4.11	2.46	6.05	4	311
15	3.30	2.21	4.90	1	436
16	3.76	2.39	5.70	1	486
17	3.60	2.47	6.09	1	351
18	4.08	2.61	6.81	1	306
19	3.35	2.50	6.26	1	345
20	3.55	2.44	5.94	1	480
21	3.69	2.61	6.79	1	443
22	3.84	2.87	8.22	1	314
23	3.03	2.45	6.01	1	361
24	4.22	3.04	9.25	1	399
25	3.89	3.02	9.11	1	452
26	4.85	3.99	15.90	1	505
27	4.44	3.17	10.03	1	416
28	4.43	3.46	11.99	1	546
29	3.95	2.93	8.60	1	469
30	5.49	3.41	11.62	1	704
31	4.53	3.31	10.95	1	513
32	5.01	3.55	12.60	1	550
33	4.61	3.53	12.47	1	450

APPENDIX A

Analysis of Utterance Length.

There were 13,719 utterances in the 33 hours of the Philippe corpus. The length of an utterance is the number of words in it, using the definition of a word as in our dictionaries. Under this definition, the mean utterance length is 3.86, and the mode length is 1. There were 3318 utterances with length 1. The longest length in the distribution was 30, and there were 2 utterances with this length. The entire distribution of utterance lengths is given in Table 28. The variance of this distribution is 7.40 (standard deviation 2.72). Pearson's skew statistic is 1.05.

As in Smith (1972), we attempted to fit the distribution of the lengths of utterances to one of several models, using simple distributions. The distributions used are the geometric (Table 29), the Poisson (Table 30), and the negative binomial (Table 31). As in previous works, the best results are obtained with the negative binomial, where the chi-square sum is 883 with 17 degrees of freedom. The geometric is second best, with a chi-square sum of 1221 with 23 degrees of freedom. The Poisson distribution has a chi-square sum of 24,386 with 10 degrees of freedom.

TABLE 28

Distribution of Utterance Lengths in Philippe, Hours 1 to 33

Length	Frequency	Sum over lengths
1	3318	3318
2	1764	5082
3	1941	7023
4	1857	8880
5	1704	10584
6	1125	11709
7	734	12443
8	493	12936
9	292	13228
10	188	13416
11	124	13540
12	53	13593
13	38	13631
14	26	13657
15	18	13675
16	15	13690
17	6	13696
18	4	13700
19	4	13704
20	3	13707
21	2	13709
22	3	13712
23	0	13712
24	2	13714
25	1	13715
26	1	13716
27	0	13716
28	1	13717
29	0	13717
30	2	13719

TABLE 29

Geometric Distribution Fit for Philippe Utterances, Hours 1-33

LENGTH X	FREQ X	THEOR.FREQ X	THEOR.PROB	CHI2
1	3318	3555.85	.26	15.91
2	1764	2634.20	.19	287.47
3	1941	1951.44	.14	.06
4	1857	1445.64	.11	117.05
5	1704	1070.94	.08	374.21
6	1125	793.37	.06	138.63
7	734	587.73	.04	36.40
8	493	435.40	.03	7.62
9	292	322.55	.02	2.89
10	188	238.94	.02	10.86
11	124	177.01	.01	15.88
12	53	131.13	.01	46.55
13	38	97.14	.01	36.01
14	26	71.96	.01	29.36
15	18	53.31	.00	23.39
16	15	39.49	.00	15.19
17	6	29.26	.00	18.49
18	4	21.67	.00	14.41
19	4	16.06	.00	9.05
20	3	11.89	.00	6.65
21	2	8.81	.00	5.27
22	3	6.53	.00	1.91
23	0	4.84	.00	
24	2	3.58	.00	4.89
25	1	2.65	.00	
26	1	1.97	.00	
27	0	1.46	.00	2.73
28	1	1.08	.00	
29	0	.80	.00	
30	2	.59	.00	.11

Geometric distribution parameter = .26
Sum of theoretical probabilities = 1.00
Chi-square sum = 1221.00
Degrees of freedom = 23

TABLE 30
Poisson Distribution Fit for Philippe Utterances, Hours 1-33

VALUES ARE SHIFTED SO THAT LENGTH 1 IS COUNTED AS 0
FOR THE DISTRIBUTION

LENGTH X	FREQ X	THEOR.FREQ X	THEOR.PROB	CHI2
1	3318	787.12	.06	8137.66
2	1764	2249.72	.16	104.87
3	1941	3215.02	.23	504.86
4	1857	3063.00	.22	474.84
5	1704	2188.63	.16	107.31
6	1125	1251.09	.09	12.71
7	734	595.97	.04	31.97
8	493	243.34	.02	256.15
9	292	86.94	.01	483.69
10	188	27.61	.00	931.78
11	124	7.89	.00	1708.43
12	53	2.05	.00	
13	38	.49	.00	
14	26	.11	.00	
15	18	.02	.00	
16	15	.00	.00	
17	6	.00	.00	
18	4	.00	.00	
19	4	.00	.00	
20	3	.00	.00	
21	2	.00	.00	
22	3	.00	.00	
23	0	.00	.00	
24	2	.00	.00	
25	1	.00	.00	
26	1	.00	.00	
27	0	.00	.00	
28	1	.00	.00	
29	0	.00	.00	
30	2	.00	.00	11631.34

Poisson distribution parameter = 2.86
Sum of theoretical probabilities = 1.00
Chi-square sum = 24385.62
Degrees of freedom = 10

TABLE 31

Negative-Binomial Distribution Fit for Philippe Utterances

Hours 1-33

PARAMETERS ARE P = .39 , R = 1.80

DISTRIBUTION IS STARTED AT X=1, AND VALUES ARE SHIFTED.

LENGTH X	FREQ X	THEOR.FREQ X	THEOR.PROB	CHI2
1	3318	2478.29	.18	284.52
2	1764	2736.35	.20	345.52
3	1941	2350.27	.17	71.27
4	1857	1826.56	.13	.51
5	1704	1344.90	.10	95.88
6	1125	957.27	.07	29.39
7	734	665.72	.05	7.00
8	493	455.19	.03	3.14
9	292	307.25	.02	.76
10	188	205.30	.01	1.46
11	124	136.06	.01	1.07
12	53	89.56	.01	14.93
13	38	58.63	.00	7.26
14	26	38.19	.00	3.89
15	18	24.77	.00	1.85
16	15	16.01	.00	.06
17	6	10.32	.00	1.81
18	4	6.63	.00	1.04
19	4	4.25	.00	
20	3	2.72	.00	.00
21	2	1.73	.00	
22	3	1.10	.00	
23	0	.70	.00	
24	2	.45	.00	
25	1	.28	.00	
26	1	.18	.00	
27	0	.11	.00	
28	1	.07	.00	
29	0	.05	.00	
30	2	.03	.00	11.29

Negative binomial distribution parameters:
p = .39 r = 1.80
Sum of theoretical probabilities = 1.00
Chi-square sum = 882.64
Degrees of freedom = 17

APPENDIX B

Analysis of Noun-phrase Length

There were 5508 noun phrases in the 9 sessions that were examined in this report. This excluded noun phrases that were in part unintelligitble. The mean length of these noun-phrases was 1.61, and the mode was 1. The variance was .85, standard deviation .92, and Pearson's skew statistic was .67.

Table 32 gives the distribution of the noun-phrase lengths. Table 33 gives the fit of the geometric distribution to the empirical distribution, Table 34 the fit of the Poisson, and Table 35 the fit of the negative binomial.

TABLE 32

Distribution of Noun-phrase Length in Noun-phrase Corpus

Length	Frequency	Cumulative frequency
1	3137	3137
2	1771	4908
3	364	5272
4	117	5389
5	87	5476
6	19	5495
7	7	5502
8	3	5505
9	1	5506
10	2	5508

TABLE 33
Geometric Distribution Fit for Philippe Noun Phrases

LENGTH X	FREQ X	THEOR.FREQ X	THEOR.PROB	CHI2
1	3137	3412.61	.62	22.26
2	1771	1298.25	.24	172.15
3	364	493.89	.09	34.16
4	117	187.89	.03	26.75
5	87	71.48	.01	3.37
6	19	27.19	.00	2.47
7	7	10.34	.00	1.08
8	3	3.94	.00	
9	1	1.50	.00	.38
10	2	.57	.00	3.59

Geometric distribution parameter = .62
Residual theoretical frequency = .35
Sum of theoretical probabilities = 1.00
Chi-square sum = 266.55
Degrees of freedom = 7

TABLE 34

Poisson Distribution Fit for Philippe Noun Phrases

VALUES ARE SHIFTED SO THAT LENGTH 1 IS COUNTED AS
0 FOR THE DISTRIBUTION

LENGTH X	FREQ X	THEOR.FREQ X	THEOR.PROB X	CHI2
1	3137	2980.78	.54	8.19
2	1771	1830.25	.33	1.92
3	364	561.90	.10	69.70
4	117	115.01	.02	.03
5	87	17.65	.00	272.40
6	19	2.17	.00	
7	7	.22	.00	
8	3	.02	.00	
9	1	.00	.00	
10	2	.00	.00	363.16

Poisson distribution parameter = .61
Residual theoretical frequency = .00
Sum of theoretical probabilities = 1.00
Chi-square sum = 715.40
Degrees of freedom = 4

TABLE 35

Negative-Binomial Distribution Fit for Philippe Noun Phrases

PARAMETERS ARE p = .72, R = 1.58

LENGTH X	FREQ X	THEOR.FREQ X	THEOR.PROB	CHI2
1	3137	3278.61	.60	6.12
2	1771	1450.01	.26	71.06
3	364	523.44	.10	48.57
4	117	174.78	.03	19.10
5	87	55.99	.01	17.17
6	19	17.48	.00	.13
7	7	5.36	.00	.50
8	3	1.62	.00	
9	1	.49	.00	
10	2	.15	.00	6.20

Negative binomial distribution parameters
p = .72   r = 1.58
Residual theoretical frequency =   .06
Sum of theoretical probabilities =   1.00
Chi-square sum = 168.91
Degrees of freedom =   5

APPENDIX C

Vocabulary of Philippe

Table 36 gives, in rank order, the words used by Philippe during the 33 sessions and their frequencies. When several words have the same frequency, the frequency is displayed only once, and the words are put together on the line with the frequency and on succeeding lines, if necessary. Table 37 gives the words in alphabetical order.

TABLE 36
Rank Ordering of the Vocabulary of Philippe

2316	est
1641	le
1531	c
1457	la
1370	&&&
1351	oui
1164	je
1137	non
1100	c*a
1079	un
1060	pas
1012	il
964	a
871	la'
860	de
777	les
771	une
653	des
639	elle
619	l
520	on
505	dans
468	a'
452	et
445	pour
444	qu

391	en
374	tu
366	fait
358	j
348	qui
345	y
334	quoi
333	ai
326	oh
311	veux
308	petit
307	sur
297	du
294	avec
287	va
277	papa
272	faire
270	voiture
262	vais
260	est-ce ou'
233	moi que
212	encore
209	regarde
204	ce madeleine maman
199	tout
192	sais
189	petite
178	mettre
177	aussi
175	ils
174	s voila'
169	pourquoi
166	comme
161	celui-la'
160	deux
159	plus
156	ah
153	fais
152	t
148	autre faut
143	monsieur
141	camion dedans
137	si
123	eau
120	comment
117	sont
115	as
114	cette mon
111	maison
110	parce train

109	eh
106	mais
105	philippe
102	voir
101	ben
98	au
97	chez se
96	bien
95	chocolat
92	bateau
91	me
90	petits
89	gros mamie
88	vois
85	moto ton voitures
84	par
83	te
82	casse' manger
79	fini
78	d
77	e'cole
76	m
75	cigarette feu
74	toi
73	rouge
72	tombe'
71	fume'e
68	croissant ma met mets
67	beaucoup peu son
66	mis
65	caca
63	ginette sa
62	chercher
61	montre
60	boum mal quand
58	ces
57	main marche
56	celle-la' musique tiens
55	puis
54	maintenant suis table
53	enlever roule
52	grosse porte trop veut
51	avait crayon elles roues tous
50	aller terre tre's
49	livre seul sucre ta
48	chien fille
47	bout noir vas
46	me'me
45	peut
44	attends plein

43	cuille're vu
42	apre's rien trois
41	ont tourne
40	e'tait lui misette tomber
39	bouton enle've monte
38	boi'te boire crayons loup marc petites trou
37	bleu broum dame garc*on mange' roue sac
36	mange souris toboggan vert
34	dessus hop tracteur
33	aux chaud dit machine peux
32	aujourd chose hui
31	cafe' chat grand joue jouer jouets yeux
30	dro'le enfants he'licopte're sous tombe viens
29	appelle donne merci ses
28	achete' briquet e'crit fleurs heure pan prendre quelle quelque soleil
27	avion be' be'be' beau brosse co'te' dur lit papier pipi te'te toutes
26	blanc croissants moteur neige rouler
25	couleur disque gens jaune perdu pied raquette toute triangle va-t verre
24	animaux bouche casse'e chaise chapeau dehors devine dire e'cris font ga'teaux hue la'-bas la'-dedans maisons ou ouvrir plai't
23	arrive belle bon e'crire monter nom queue
22	alors bruit carre' chaussons fene'tre mai'tresse morceau peau tourner
21	bras e'chelle envers grande ici passe pris route
20	autres boule es lapin mane'ge ouvre parler prends tailler tient tombe'e
19	bois chambre chats clara euh la'-haut partie pistolet rails rue sens seule vient vrai
18	arbre boutons chaussures fermer fil lumie're ouais tortue truc ve'lo
17	avant bateaux bonhomme casse choses connais coupe crois fleur gentil haut lune ne personne poisson tire travailler vitre
16	aiguille alle' biberon bol cheval de'ja' feuilles froid garage hein jouet mains montagne place toit tour trous ventre voit
15	aime ballons bou cage chemine'e clef cuille'res de'fait docteur donner fabrique girafe jardin josiane pain parti peigne pieds pre't sort sortir trucs vaches
14	balle cigarettes cochon enleve' lait magne'tophone marcher me'tro mouchoir mouton orange pelle peut-e'tre pont pot quel sie'ge vite
13	allez attention bouge boules bru'ler ciel de'de' derrie're dessous e'cureuil gentille he' me'dicament montrer navion nuit peur poissons

portes poule poumandarinette rentre rond rouges
singe stylo tiroir vent volant vous
12 accrocher cherche cour doigt doigts donne'
e'couter ferme' ge'rard i langogne madame marron
miam mien nez partir ramasse rateau regarder
rentrer sable tableau tartine te'le'phone wagon
11 appuie bonbons bu cache' casser chaque che'vre
collier con coup course couteau dents difficile
fauteuil gronomme joue' jour leur leurs locomotive
malade marteau me'chant michel monte' n nen parle
presque re'parer sage tambour trains tut

10 accident acheter arrive' asseoir avais baisse bas
be'te bonne canard cendrier chai'ne cine'ma
confiture droit ferme ficelle gare gauche gris
he'lice joli langue livres longtemps lunettes
miettes monde moutons myriam ours pascal poubelle
poupou prend ranger renard rou sert souviens tas
vieux vole

9 ai'e allumette ame'ne bague ballon besoin boi'tes
bouts cache caravane celui ceux-la' chausson colle
cou couper cours culotte de' de'faire dis envole
faisait fois front ga'teau gomme grosses guitare
idiot jambes jeter la'-dessus laisse' lancer
lequel lire long mer ouvert photo photos pipe
poche puzzle remorque roum sucres taureau
te'le'vision tes touche travers trouve' vis vont

8 accroche ache'te allumer attrape balaie
be'tonneuse boulogne chasser chiens coffre
conduire content cuiller dames dessine dessins
doit farine fene'tres fou gant garc*ons genoux
grue histoire ide'e jaunes journal loin lourd
manque marco mauvais me'dicaments oiseaux orsay
pantalon papillon pauvre plantes porter poupe'e
promener quatre ramasser re'pare' remettre reste
rogalski ronds rose salon serpent tige toujours
traces travail ventilateur verte violet
voiture-la'

7 allumettes arbres attraper au-revoir ba'ton boirat
bols bru'le' bulles cet cheveux clic conduiteur
dessin dessine' disques donne-moi fourchette halle
herbe limoges losange lumie'res manteau mienne
monsieurs moteurs oiseau oreilles ouah oublie' ouh
outils ouvertes panier passer piano pique pleure'
pneu police pompier poussie're rester restes roses
roule' saute sorti tigre tonnerre trottoir ve'los
vins zut

6 accroche' alle's animal annulaire appareil
aristochats armoire arre'te arre'ter attendre
balai battre belles benne beuh blancs bleus bosse
bru'le bru'le'e cache-cache casque casserole
cochons corbeille court couvercle cuisine da
dangereux de'fais dort enfant escalier failli
figure froide fume garde groseilles h heures janou
jeu jolie jure jus kangourou le'gos lion loups
marine martine me'tal merde mes micro mieux
moitie' monsieur-la' morceaux mouille'e oeuf
oreille oui'e page papiers pardon pareil partout
peigner piquer plancher plusieurs poi pointe pomme
porte' porte-monnaie prome'ne quelqu racheter
range refaire refais remets retrouve' sait sans
seguin serviette soye tabouret temps tirer tousse
travaille trouve venir veulent vienne vilaine vin
vitres

5 abat-jour accroches allume amener amusant apporter
araigne'e assiette assis attaquer au-dessus avions
banquette bats be'ne'dicte be'tise biscotte
blanches bleues boi'te-la' bouteilles cacher
cadeau camions car carrelage cha'teau chante
chauffe-eau classe clefs cogner coin colorier
conduit couche' de'chire de'chirer demander
demi-soleil descendre descends devant devient
discutent drapeau dro'les e'coute e'glise
e'le'phants e'ponge e'tais ennuis envoie essuyer
eu fasses filles fourrure gre've he'risson hi
laine lampe lave laver leve'e luge magasin miette
minou montres oeufs ouverte ouvres panneau parfois
part pattes pauvres pluie possible pousse re'pare
reconnai't regarde' regardez rentre' retourne
retourner revient rigolo robe ronde salir seau
sent serre seulement souffler sourcils tabe'
talle' tapis timbale tombent tord tourne-vis
tousser tuyau use' vache valise valises ve'tement
verres vraie wagons zorro

4 a'ne accroche'e aimant ajourabe alle'es allemagne
arcs-en-ciel arre'tez-vous assieds aura auras
autos bande barres be'te-la' beu billes biscottes
bitonniau blanquette bobo bolle' bouteille
bouton-la' brouette cabines cabriole cache-le
camion-la' casquette casse'es celle chemise chevau
chevaux cinq clair cogne' corde couverture
croise'e croit crou'te debout demi-rond dormir
drapeaux droite e'craser e'le'phant e'lectrophone
ennuie ensemble envie envoyer essaie essaye eux
fesse'e filet fle'che fondu fort fous frigidaire

fromage fumer gag glisse glisser gourmand goutte
hexagone hier hoquet huit importe jumelles laisse
lapins lavabo le'go leu lions lis machines
magne'ophone malheur malo masque matin me'canique
mets-le mickey montbrun montparnasse montre-moi
motos moudre moulin moyen mum mur musiques
myrtilles nenfant nettoie nicole noire noirs nu
nus panne papillons paquet passage penses-tu peuh
philou pistolets plouf pompiers poules poum
pourrais poussie'res proble'mes pu puisque
puzzle-la' re'veil recasse' regardent rentre'e
renverse' rhabiller rit robot rocher roulent sales
salete's salit samedi scie scoch se'che skis
soigne suite syndicat tableaux taille tape tase
te'le'phe'rique tien tienne toboggans toucher
tracteurs tuer ve'tements venu vide vider vieille
vilain voudrais voulez vraies zoo

3 achete's ailleurs alle'e allo amuse angleterre
antennes appellent appelles appuies asseoit
attache attacher aurait autour bab babar baguette
bain balanc*oire balancer bandes bates be'tises
beurre bigoudis blanche bleue bonjour bottes
bouchon boudie bougie bracelet brou bulle cadeaux
canards canne caravanes carnet cartable casse's
ce' chameau champignon chanson chaperon charriots
che'vres chemine'es cirage cire' clown clowns
coincent colique commerce complique' compresseur
cor courir cravate creux crip crochet crocodile
croyais cygne danielle de'bouche' de'jeuner
de'moli de'monte' de'montes de's de'sole' demain
demande dernier descend dessiner dimanche disent
dommage dos drabon dure e'teindre e'teins e'toiles
e'tonne e'tre eiffel embouteillage emme'ne emmener
emporte's endroit enleve's escalou essayer est-c*a
expliquer expre's faite fe'te film fin finir finis
fleur-la' fruits gages galipette galoper garder
gare'es garer genou glace glaces glisse' gorge
grands habille habille' habite hamster hum
isabelle jeanine jette jouerait judo lacets
laquelle le' le've lourde maison-la' malades
manges marque' masou me'me're menteur meubles
miaou microbes mie' migrec mireille mise'ricordes
morde mordes mort morts mouchelette mouches
mouille' nager nettoyer neuf noiraud nous ombre pa
par-dessus parc parents parking pars partent
passe' patin patrick pe'tales peinture perdre
phare pie'ce pinceau piqu're pis piste plafond
plage plastique plateau pleure pleut plier pneus
poltron porquerolles pose poser pouce poupe'es

premie're premier prendra puzzles-la' ramasse' ramasses range' rate recasser reconnu remet remonte renverse repartir ressemblent ressort retombe rhabillette roue-la' rouleau rouleze routes sale sar saute' semblant sera soigne' solide souffleu soupe stylos tache' tape' te'tes tenir tiroirs toilette tonne tonneau tortille touche' tourterelle tourterelles transporte travaille' travaillent travailles triangles trompette troup truc-la' veaux verts veux-tu voile volent voler vraiment vue yaourt zizi zoos zoutils

2 abord accidents accrochent aiguilles air allume' amusent ancre ani anneaux appartement apprend appuyer argent arre'te' arre'te'e arre'tes arre'tez assez assi-les-moulineaux attend attendes attrape' autocar avaient avoir baguitte baisers baisse' banc barbe barrie're barrie'res base be'leu be'tement beguitte biche biches bill bille billet bizarre blagueur ble' bobine boiboiboiboi boit bolo bons bord boue'e bouette bougies boup bout-la' boutiques brioche brosser bru'lent bussy cacahue'te cache'e calembour carole carqueranne carre'es carte casserait cerises chaises chamois champignons champs chapeaux charriot cherches chevals chiffon chigne circulation ciseaux cloche co'te's coince colle' coller combien commissions comprends compte compte's construction construit contents coquin corne couche coud coudes coule couleurs coupe' coups courant coussin coussins couture crasseux cravates crayon-la' cre'pe crocrocro croq croquer crotte' cube dada de'barbouiller de'but de'chire' de'fais-le de'fendre de'jeune' de'molir de'moliront de'monte'e demassieu demeussieu dernie're dessines dirait dirige dis-donc dis-moi donnes dors dossier doux drap du' e e' e'clant e'clater e'crise e'glique e'lectrique e'pe'e e'tage e'tage-la' e'tages e'te' e'tiquette e'tonnant embe'tes emporte encre enle'ves enleve'e entend entends entier envole'e envoye' estatues existe fabrique' fabriquer faim faisaient fatigue fauche faute ferme'e ferment fesses flu'te flu'tes fourneau fourreau foutre fronde frotter fusil fusils gagner gare' gilet gneu gou't grand-me're grandes grenouille grimper gueule guidons guiges ho'tel image indien installe installer jamais jambe jean-michel jeu-la' jeudi jeuh jusqu klaxon laisse-moi lance'e leuleu leve' lieu lise longue lourds luxembourg machin machine-la' magnifique

mai's mangeaient mangent manifestation manque'
mapa marchand marche' margoulette marionnettes
marrant mathalie me'la me'mes menton messieurs
meuh michambe migre milieu minouche mithalie
montrerai montsouris moque mordu moto-la' mots
mouche moulineaux nage nanou narbre nautres no'tre
noie nours nouveau nouvelle nuage nue ocarine oeil
offre otites oubliais ouvre-moi ovale pages pam
paquets parapluie parties partiras passe'e patins
patte pe'se pense' perd peuvent pierres piquants
planques plat plates pleine pleurer plume pof
poids points pointu poire portent postale pouf
pourrait pousser poussin poussins pouze pre'pare
prendait pull-over quille quilles raccourcie
radiateur raisin raser rati rats re'fle'chis
re'galer realle's redorme reenle've refroidisse
remis remonter remue rends renverser reste'
retombe' retourer reviendra reviens rhinoce'ros
rivie're ro rochers rondes roues-la' roulettes
rououououou samedie' samos sante' sauver savoir
savon se'que' secoue semaine serait servir sieste
six ski soif sommeil sors sorte sortie sortira sou
souffle suffit surtout tache'e taille-crayon tant
tasse tata taxi te'le'phoner te'le'visions teau
ter tiennent tire' tire-bouchon tirelire tissus
toffe tonneaux tordue tournent toux trai'ne'
traire tramway tremper triches triste trompe
trompe' trucs-la' tue vacances veau vend ver vie
vieilles viendront ville voi voisin voulais voulu
vrais vroum zailes zoiseau

1 abi'me'e abricure accompagne'e accorde'on achete'e
achete'es ae'rogare agne's agrafe agrafe's agrafer
aides aie'e aigle aiguille-la' ailes aillent ainsi
aiou alles alligator allons allume'e ambulances
ami amis amour amuser ancien anglais antenne
aouiii ape'ritif appele' apporte' apporte's
apprendre appuyant arc-en-ciel arranger arre'tent
arre'tera arriver arriveront arroser assieds-toi
assiettes attache' attaque'e attendais attende
attrapait attrape-le auront auto-la' autobus
autocars autoroute avale avancer avis aye ayent ba
bababababa babi baboum bagnole baguette-la' bah
baigne' baigner baille baiser baisse'e baisser
balais balance balaye'es balcon balles ballon-la'
bambi bancs barreau barreaux barres-la' bassin bat
bateau-la' batter bave be'be's be'e' be'ta
be'tes-la' be'toleu be'toleuse be'ton bec beleu
bibibibibobodidabe bibibobabibi bibiboboline
bibibobolune biento't blablablabla blesse' blonds

boh boivent bol-la' bombe bouboubou bouche'e
boucherie bouh bouillante bouloum boulu bouscule
bousse bouteille-la' boutons-la' bouts-la'
bracelets branches brosses bruit-la' brun bureau
burode c* ca'liner cabane cabinet cacahue'tes cace
cacheras caches cages cahier cailloux caisse
campagne campagnes camping caresses carnassier
carnassiers carre'e carreau cartes casquettes cass
casseroles castor ce'ce's celles celles-la' cendre
cerf cerise cervelle chaleureux chambre-la'
chandelle chansons chante' chantent chanter
chapeau-la' chasse chasse' chasseur chatouille
chaude chauds chauffer chaussettes che che'vre-la'
cherchais chercherai cherchons cheveu chichis
chien-la' chinois chonchon chonchons chose-la'
chou chouette cigares cire claque cloute'
co'te'-la' codile coge' cognac cogne'e coince'
colle'e collier-la' coloriages compote confitures
connaissais construire contente copains coqueluche
coquillages cordes cornichon couche'e coucher
coude coudre coule' couler couloir coupe'e coupe's
coupent coupes courait couru couteaux couverts
couvre craie cre'me cre'p cre'pes creux-la' creve'
crier cro crochets croi croisse crole crooooh
croque' crou'te-la' cuille're-la' cuiller-la'
cuire cuisse cuisses dame-la' dansent
de'barbouille de'coller de'coupe' de'croche'
de'fais-moi de'faits de'gringole' de'molira
de'monter de'passe de'passent de'pe'che de'raille
de'shabille de'shabille' de'tache'e de'visse dent
dentifrice depuis descende descendu descendue
dessin-la' dessine'e devais dieu diras disputent
dix dix-neuf dodo doivent domdomdomdom donnait
donne'es donne-le dore' dormant dormi douze dragon
dro'lement droi e'c e'clabousse e'clair e'coute'
e'coutes e'couteur e'crou e'dicaments e'e'h
e'lastique e'lastiques e'lectriques e'pouvantail
e'taient e'tamines e'teint e'tudiants eau-la'
edelweiss efface' ele'phant embe'tant embe'te
emporte' emporter endort enfants-la' enferme
enferme' enlevant enregistre entiers entre
envoie-le es-tu escaliers escoulou espe'ce
essayait essuie eue existe' explique fabouilles
fac*on face facteur facture fais-le fais-moi
faisais faisant faites famille-la' fan fanfare
fascine' fasse fatigant fatigue' faucille
fauteuils faux ferai ferait fermais ferme's
fesse'es festation feu-la' feuille feux
ficelle-la' fifille fille-la' fils flamme flic
flip flotte foie foievelle fond forme fragile

franc*ois fre're freine freine' froisse' fuirait
fumes funiculaire fuse'e ga ga'teau-la' gagne'
gants garage-la' garc*on-la' gaz ge'nait ge'nant
ge'ne genre gentils gentiment glace'e gne' gomment
goulue gourmande gouttes grain grains gratte
grenouilles griffes grille grimpe grog gronde'
grossi gruye're gue'ri gue'rie guidons-la' guilis
guy gyroscope habiller habitent hache
he'licopte'res helena hibous hippocampe houp house
huma hyppolite iaa idiots index ion ira irait iras
janoucha japonaise jean-paul jet jete' jetons
jettent jetterai jouaient jouais jouent jouera
jouerai journe'e juste kira knickers kola la'r
labo laid laids laisse'e laisse'es laisser laisses
lala lance lave' le'cher le've-toi lettre lettres
lever li lilalalala lili liliane loine lolo longs
loze're luges lunette lunettes-la' magasin-la'
mai'tresses majeur mama mamaman man mange'-cale
mange-cale mangeait mangeraient mani manou
manquait manquer manteau-la' marchent
marcominouche mari marianne marmotte marraine
marrons marseille masque-la' me'chants me'got
me'me'me'seu me'seu mesdames mets-les mettant
mette mettes mettra mettrai meuble meueu miens
mimi mimimorire minet mise mitalie mobi mobilis
moche moches modeur mollets moment monse montagnes
montaient montbrin monte's montent montes
montre-la monument morceau-la' morte mot
moteur-la' mouche' mouchelettes mouds mouflon
mouille's mouiller moules moulins moulu moustache
moustaches mur-la' murs muse' muse'se'se'
muse'se'se'seau muse'se'seau muse'seau museau
musicien na nageoires nallumettes nanimals navions
nettoye' neus neuse ni nid noe'l noiseau noix
normal noye' nulle nume'ros o oblige's obstine
occupe oeils oignons omm ope're' oreiller otite
oublie oublie'e ouf ouie' ouverts ouvrai pa'te paf
paff panneaux panpan pansement papas papiers-la'
parai't parfait paris parles partage partira
partis pastroui'llarare pastroui'llare'
pastrouillai'ra pastrouillai're' pastrouilles
pataboum pe pe'che pe'cheux pe'dale peignes peine
penche pendant penses perdue perroquet perroquets
personnes petit-de'jeuner phares pie'tons pierre
pierrot pigeon pin pingouin pinocchio pique'
piquent placard plaf plaisantin planche plas plie
plof poche-la' pochette pois pomp porte-moi
portie're pose-les positoire poubelles poulet
pourra pourront pouzle pre'au pre's pre'te-moi
prende prochaine profondeur prote'ger publicite'

puisqu pull quelles quels raconte raconter
racontes radio radotes raisins raison ramasse'e
rames rappelle rapporte' ras rat rateaux rattrape
rattrape' re' re'chauffer re'fle'chit re'pare'e
re'ussi re'veille rec*u recacheras recoller
reconnais reconstruit recoucher rectangulaire
reculotte rede'monter refait refermer refermes
refroidi regardes regardons relaver releve'e
remets-la remettez-les remmene' remplacer rentrera
rentrerait renverse'es renversent renvoyer
repartie ressortir ressorts restaurant restent
resteras retrouve retrouver retrouvera revenue
revoila' rire risquais riz robert robes robinet
robinets roche rondelles roros saint-cloud sali
salle salut sandales sang saperlipopette sapin
sautent sauter sauve savent scania sciure se'chent
se'que's second seize sel sens-la' serai seront
servent serviette-la' serviettes shampooing
sifflet singes sinon sire'ne sirop so soigner
soignera soignerait soir sonne' sophie sortant
sortirai souffe soule've souleve' sourires souvent
soyent spirou statue su sucre' super suppositoire
suppositoires surprise symbolique tache taintain
te'gon te'le'phone' temps-la' tenait tente-la'
the' tic-tac ticket tickets timbales tinette tires
to't toits tom tombe'es tombe's tomberas tordu
tordues tortues touf touille toupie tour-la'
tousses trai'ner train-la' tranquille transforme'
transparente treize trempe trempes troisie'me
trombone tronc trotinette trotinettes troue's
trous-la' trouve'e trouve's trouver tube tunnel
tuser tuyaux tyran u usine vague ve' venue venus
verrai verras vers verserai vert-la' vertes
verture veste viendra viendras viennent village
violettes violon visse visse-le visser vit
voitures-la' vole' vomi vomir voudrait voyons
wagon-la' xagone zarrivent ze'coutent zencore
zigzag zont 0h 10

Types = 2698
Tokens = 56,982

TABLE 37
Alphabetical Ordering of the Vocabulary of Philippe

a	964	a'	468	a'ne	4
abat-jour	5	abi'me'e	1	abord	2
abricure	1	accident	10	accidents	2
accompagne'e	1	accorde'on	1	accroche	8
accroche'	6	accroche'e	4	accrochent	2
accrocher	12	accroches	5	ache'te	8
achete'	28	achete'e	1	achete'es	1
achete's	3	acheter	10	ae'rogare	1
agne's	1	agrafe	1	agrafe's	1
agrafer	1	ah	156	ai	333
ai'e	9	aides	1	aie'e	1
aigle	1	aiguille	16	aiguille-la'	1
aiguilles	2	ailes	1	aillent	1
ailleurs	3	aimant	4	aime	15
ainsi	1	aiou	1	air	2
ajourabe	4	alle'	16	alle'e	3
alle'es	4	alle's	6	allemagne	4
aller	50	alles	1	allez	13
alligator	1	allo	3	allons	1
allume	5	allume'	2	allume'e	1
allumer	8	allumette	9	allumettes	7
alors	22	ambulances	1	ame'ne	9
amener	5	ami	1	amis	1
amour	1	amusant	5	amuse	3
amusent	2	amuser	1	ancien	1
ancre	2	anglais	1	angleterre	3
ani	2	animal	6	animaux	24
anneaux	2	annulaire	6	antenne	1
antennes	3	aouiii	1	ape'ritif	1
appareil	6	appartement	2	appele'	1
appelle	29	appellent	3	appelles	3
apporte'	1	apporte's	1	apporter	5
apprend	2	apprendre	1	appuie	11
appuies	3	appuyant	1	appuyer	2
apre's	42	araigne'e	5	arbre	18
arbres	7	arc-en-ciel	1	arcs-en-ciel	4
argent	2	aristochats	6	armoire	6
arranger	1	arre'te	6	arre'te'	2
arre'te'e	2	arre'tent	1	arre'ter	6
arre'tera	1	arre'tes	2	arre'tez	2
arre'tez-vous	4	arrive	23	arrive'	10
arriver	1	arriveront	1	arroser	1
as	115	asseoir	10	asseoit	3

assez 2
assi-les-moulineaux 2
assieds 4
assieds-toi 1
assiette 5
assiettes 1
assis 5
attache 3
attache' 1
attacher 3
attaque'e 1
attaquer 5
attend 2
attendais 1
attende 1
attendes 2
attendre 6
attends 44
attention 13
attrapait 1
attrape 8
attrape' 2
attrape-le 1
attraper 7
au 98
au-dessus 5
au-revoir 7
aujourd 32
aura 4
aurait 3
auras 4
auront 1
aussi 177
auto-la' 1
autobus 1
autocar 2
autocars 1
autoroute 1
autos 4
autour 3
autre 148
autres 20
aux 33
avaient 2
avais 10
avait 51
avale 1
avancer 1
avant 17
avec 294
avion 27
avions 5
avis 1
avoir 2
aye 1
ayent 1
ba 1
ba'ton 7
bab 3
bababababa 1
babar 3
babi 1
baboum 1
bagnole 1
bague 9
baguette 3
baguette-la' 1
baguitte 2
bah 1
baigne' 1
baigner 1
baille 1
bain 3
baiser 1
baisers 2
baisse 10
baisse' 2
baisse'e 1
baisser 1
balai 6
balaie 8
balais 1
balanc*oire 3
balance 1
balancer 3
balaye'es 1
balcon 1
balle 14
balles 1
ballon 9
ballon-la' 1
ballons 15
bambi 1
banc 2
bancs 1
bande 4
bandes 3
banquette 5
barbe 2
barreau 1
barreaux 1
barres 4
barres-la' 1
barrie're 2
barrie'res 2
bas 10
base 2
bassin 1
bat 1
bate 5
bateau 92
bateau-la' 1
bateaux 17
bates 3
batter 1
battre 6
bave 1
be' 27
be'be' 27
be'be's 1
be'e' 1
be'leu 2
be'ne'dicte 5
be'ta 1
be'te 10
be'te-la' 4
be'tement 2
be'tes-la' 1
be'tise 5
be'tises 3
be'toleu 1
be'toleuse 1
be'ton 1
be'tonneuse 8
beau 27
beaucoup 67
bec 1
beguitte 2
beleu 1
belle 23
belles 6
ben 101
benne 6
besoin 9
beu 4
beuh 6
beurre 3
biberon 16
bibibibibobodidabe 1

bibibobababibi 1
bibiboboline 1
bibibobolune 1
biche 2
biches 2
bien 96
biento't 1
bigoudis 3
bill 2
bille 2
billes 4
billet 2
biscotte 5
biscottes 4
bitonniau 4
bizarre 2
blablablabla 1
blagueur 2
blanc 26
blanche 3
blanches 5
blancs 6
blanquette 4
ble' 2
blesse' 1
bleu 37
bleue 3
bleues 5
bleus 6
blonds 1
bobine 2
bobo 4
boh 1
boi'te 38
boi'te-la' 5
boi'tes 9
boiboiboiboi 2
boirat 7
boire 38
bois 19
boit 2
boivent 1
bol 16
bol-la' 1
bolle' 4
bolo 2
bols 7
bombe 1
bon 23
bonbons 11
bonhomme 17
bonjour 3
bonne 10
bons 2
bord 2
bosse 6
bottes 3
bou 15
bouboubou 1
bouche 24
bouche'e 1
boucherie 1
bouchon 3
boudie 3
boue'e 2
bouette 2
bouge 13
bougie 3
bougies 2
bouh 1
bouillante 1
boule 20
boules 13
boulogne 8
bouloum 1
boulu 1
boum 60
boup 2
bouscule 1
bousse 1
bout 47
bout-la' 2
bouteille 4
bouteille-la' 1
bouteilles 5
boutiques 2
bouton 39
bouton-la' 4
boutons 18
boutons-la' 1
bouts 9
bouts-la' 1
bracelet 3
bracelets 1
branches 1
bras 21
brioche 2
briquet 28
brosse 27
brosser 2
brosses 1
brou 3
brouette 4
broum 37
bru'le 6
bru'le' 7
bru'le'e 6
bru'lent 2
bru'ler 13
bruit 22
bruit-la' 1
brun 1
bu 11
bulle 3
bulles 7
bureau 1
burode 1
bussy 2
c 1531
c* 1
c*a 1100
ca'liner 1
cabane 1
cabines 4
cabinet 1
cabriole 4
caca 65
cacahue'te 2
cacahue'tes 1
cace 1
cache 9
cache' 11
cache'e 2
cache-cache 6
cache-le [illegible]
cacher 5
cacheras 1
caches 1
cadeau 5
cadeaux 3
cafe' 31
cage 15
cages 1
cahier 1
cailloux 1

caisse 1
calembour
camion 141
camion-la' 4
camions 5
campagne 1
campagnes 1
camping 1
canard 10
canards 3
canne 3
car 5
caravane 9
caravanes 3
caresses 1
carnassier 1
carnassiers 1
carnet 3
carole 2
carqueranne 2
carre' 22
carre'e 1
carre'es 2
carreau 1
carrelage 5
cartable 3
carte 2
cartes 1
casque 6
casquette 4
casquettes 1
cass 1
casse 17
casse' 82
casse'e 24
casse'es 4
casse's 3
casser 11
casserait 2
casserole 6
casseroles 1
castor 1
ce 204
ce' 3
ce'ce's 1
celle 4
celle-la' 56
celles 1
celles-la' 1
celui 9
celui-la' 161
cendre 1
cendrier 10
cerf 1
cerise 1
cerises 2
cervelle 1
ces 58
cet 7
cette 114
ceux-la' 9
cha'teau 5
chai'ne 10
chaise 24
chaises 2
chaleureux 1
chambre 19
chambre-la' 1
chameau 3
chamois 2
champignon 3
champignons 2
champs 2
chandelle 1
chanson 3
chansons 1
chante 5
chante' 1
chantent 1
chanter 1
chapeau 24
chapeau-la' 1
chapeaux 2
chaperon 3
chaque 11
charriot 2
charriots 3
chasse 1
chasse' 1
chasser 8
chasseur 1
chat 31
chatouille 1
chats 19
chaud 33
chaude 1
chauds 1
chauffe-eau 5
chauffer 1
chaussettes 1
chausson 9
chaussons 22
chaussures 18
che 1
che'vre 11
che'vre-la' 1
che'vres 3
chemine'e 15
chemine'es 3
chemise 4
cherchais 1
cherche 12
chercher 62
chercherai 1
cherches 2
cherchons 1
cheval 16
chevals 2
chevau 4
chevaux 4
cheveu 1
cheveux 7
chez 97
chichis 1
chien 48
chien-la' 1
chiens 8
chiffon 2
chigne 2
chinois 1
chocolat 95
chonchon 1
chonchons 1
chose 32
chose-la' 1
choses 17
chou 1
chouette 1
ciel 13
cigares 1
cigarette 75
cigarettes 14
cine'ma 10
cinq 4
cirage 3
circulation 2
cire 1
cire' 3
ciseaux 2
clair 4

claque 1	clara 19	classe 5
clef 15	clefs 5	clic 7
cloche 2	cloute' 1	clown 3
clowns 3	co'te' 27	co'te'-la' 1
co'te's 2	cochon 14	cochons 6
codile 1	coffre 8	coge' 1
cognac 1	cogne' 4	cogne'e 1
cogner 5	coin 5	coince 2
coince' 1	coincent 3	colique 3
colle 9	colle' 2	colle'e 1
coller 2	collier 11	collier-la' 1
coloriages 1	colorier 5	combien 2
comme 166	comment 120	commerce 3
commissions 2	complique' 3	compote 1
comprends 2	compresseur 3	compte 2
compte's 2	con 11	conduire 8
conduit 5	conduiteur 7	confiture 10
confitures 1	connais 17	connaissais 1
construction 2	construire 1	construit 2
content 8	contente 1	contents 2
copains 1	coqueluche 1	coquillages 1
coquin 2	cor 3	corbeille 6
corde 4	cordes 1	corne 2
cornichon 1	cou 9	couche 2
couche' 5	couche'e 1	coucher 1
coud 2	coude 1	coudes 2
coudre 1	coule 2	coule' 1
couler 1	couleur 25	couleurs 2
couloir 1	coup 11	coupe 17
coupe' 2	coupe'e 1	coupe's 1
coupent 1	couper 9	coupes 1
coups 2	cour 12	courait 1
courant 2	courir 3	cours 9
course 11	court 6	couru 1
coussin 2	coussins 2	couteau 11
couteaux 1	couture 2	couvercle 6
couverts 1	couverture 4	couvre 1
craie 1	crasseux 2	cravate 3
cravates 2	crayon 51	crayon-la' 2
crayons 38	cre'me 1	cre'p 1
cre'pe 2	cre'pes 1	creux 3
creux-la' 1	creve' 1	crier 1
crip 3	cro 1	crochet 3
crochets 1	crocodile 3	crocrocro 2
croi 1	crois 17	croise'e 4
croissant 68	croissants 26	croisse 1
croit 4	crole 1	crooooh 1
croq 2	croque' 1	croquer 2
crotte' 2	crou'te 4	crou'te-la' 1
croyais 3	cube 2	cuille're 43

cuille're-la' 1
cuiller-la' 1
cuisse 1
cygne 3
d 78
dame 37
dangereux 6
dansent 1
de'barbouille 1
de'but 2
de'chirer 5
de'croche' 1
de'fais 6
de'fait 15
de'gringole' 1
de'jeuner 3
de'molira 1
de'monte'e 2
de'passe 1
de'raille 1
de'shabille' 1
de'visse 1
dehors 24
demander 5
demi-rond 4
dentifrice 1
dernie're 2
des 653
descendre 5
descendue 1
dessine 8
dessiner 3
dessous 13
devais 1
devine 24
dimanche 3
dire 24
dis-donc 2
disent 3
disques 7
dix-neuf 1
doigt 12
doivent 1
donnait 1
donne'es 1
donner 15
dormant 1
dors 2
dossier 2
drabon 3

cuille'res 15
cuire 1
cuisses 1

da 6
dame-la' 1
danielle 3
de 860
de'barbouiller 2
de'chire 5
de'coller 1
de'de' 13
de'fais-le 2
de'faits 1
de'ja' 16
de'moli 3
de'moliront 2
de'monter 1
de'passent 1
de's 3
de'sole' 3
debout 4
demain 3
demassieu 2
demi-soleil 5
dents 11
dernier 3
descend 3
descends 5
dessin 7
dessine' 7
dessines 2
dessus 34
devant 5
dieu 1
dirait 2
dirige 2
dis-moi 2
disputent 1
dit 33
docteur 15
doigts 12
domdomdomdom 1
donne 29
donne-le 1
donnes 2
dormi 1
dort 6
doux 2
dragon 1

cuiller 8
cuisine 6
culotte 9

dada 2
dames 8
dans 505
de' 9
de'bouche' 3
de'chire' 2
de'coupe' 1
de'faire 9
de'fais-moi 1
de'fendre 2
de'jeune' 2
de'molir 2
de'monte' 3
de'montes 3
de'pe'che 1
de'shabille 1
de'tache'e 1
dedans 141
demande 3
demeussieu 2
dent 1
depuis 1
derrie're 13
descende 1
descendu 1
dessin-la' 1
dessine'e 1
dessins 8
deux 160
devient 5
difficile 11
diras 1
dis 9
discutent 5
disque 25
dix 1
dodo 1
doit 8
dommage 3
donne' 12
donne-moi 7
dore' 1
dormir 4
dos 3
douze 1
drap 2

drapeau 5	drapeaux 4	dro'le 30
dro'lement 1	dro'les 5	droi 1
droit 10	droite 4	du 297
du' 2	dur 27	dure 3
e 2	e' 2	e'c 1
e'chelle 21	e'clabousse 1	e'clair 1
e'clant 2	e'clater 2	e'cole 77
e'coute 5	e'coute' 1	e'couter 12
e'coutes 1	e'couteur 1	e'craser 4
e'crire 23	e'cris 24	e'crise 2
e'crit 28	e'crou 1	e'cureuil 13
e'dicaments 1	e'e'h 1	e'glique 2
e'glise 5	e'lastique 1	e'lastiques 1
e'le'phant 4	e'le'phants 5	e'lectrique 2
e'lectriques 1	e'lectrophone 4	e'pe'e 2
e'ponge 5	e'pouvantail 1	e'tage 2
e'tage-la' 2	e'tages 2	e'taient 1
e'tais 5	e'tait 40	e'tamines 1
e'te' 2	e'teindre 3	e'teins 3
e'teint 1	e'tiquette 2	e'toiles 3
e'tonnant 2	e'tonne 3	e'tre 3
e'tudiants 1	eau 123	eau-la' 1
edelweiss 1	efface' 1	eh 109
eiffel 3	ele'phant 1	elle 639
elles 51	embe'tant 1	embe'te 1
embe'tes 2	embouteillage 3	emme'ne 3
emmener 3	emporte 2	emporte' 1
emporte's 3	emporter 1	en 391
encore 212	encre 2	endort 1
endroit 3	enfant 6	enfants 30
enfants-la' 1	enferme 1	enferme' 1
enle've 39	enle'ves 2	enlevant 1
enleve' 14	enleve'e 2	enleve's 3
enlever 53	ennuie 4	ennuis 5
enregistre 1	ensemble 4	entend 2
entends 2	entier 2	entiers 1
entre 1	envers 21	envie 4
envoie 5	envoie-le 1	envole 9
envole'e 2	envoye' 2	envoyer 4
es 20	es-tu 1	escalier 6
escaliers 1	escalou 3	escoulou 1
espe'ce 1	essaie 4	essayait 1
essaye 4	essayer 3	essuie 1
essuyer 5	est 2316	est-c*a 3
est-ce 260	estatues 2	et 452
eu 5	eue 1	euh 19
eux 4	existe 2	existe' 1
explique 1	expliquer 3	expre's 3
fabouilles 1	fabrique 15	fabrique' 2
fabriquer 2	fac*on 1	face 1

facteur 1	facture 1	failli 6
faim 2	faire 272	fais 153
fais-le 1	fais-moi 1	faisaient 2
faisais 1	faisait 9	faisant 1
fait 366	faite 3	faites 1
famille-la' 1	fan 1	fanfare 1
farine 8	fascine' 1	fasse 1
fasses 5	fatigant 1	fatigue 2
fatigue' 1	fauche 2	faucille 1
faut 148	faute 2	fauteuil 11
fauteuils 1	faux 1	fe'te 3
fene'tre 22	fene'tres 8	ferai 1
ferait 1	fermais 1	ferme 10
ferme' 12	ferme'e 2	ferme's 1
ferment 2	fermer 18	fesse'e 4
fesse'es 1	fesses 2	festation 1
feu 75	feu-la' 1	feuille 1
feuilles 16	feux 1	ficelle 10
ficelle-la' 1	fifille 1	figure 6
fil 18	filet 4	fille 48
fille-la' 1	filles 5	film 3
fils 1	fin 3	fini 79
finir 3	finis 3	flamme 1
fle'che 4	fleur 17	fleur-la' 3
fleurs 28	flic 1	flip 1
flotte 1	flu'te 2	flu'tes 2
foie 1	foievelle 1	fois 9
fond 1	fondu 4	font 24
forme 1	fort 4	fou 8
fourchette 7	fourneau 2	fourreau 2
fourrure 5	fous 4	foutre 2
fragile 1	franc*ois 1	fre're 1
freine 1	freine' 1	frigidaire 4
froid 16	froide 6	froisse' 1
fromage 4	fronde 2	front 9
frotter 2	fruits 3	fuirait 1
fume 6	fume'e 71	fumer 4
fumes 1	funiculaire 1	fuse'e 1
fusil 2	fusils 2	
ga 1	ga'teau 9	ga'teau-la' 1
ga'teaux 24	gag 4	gages 3
gagne' 1	gagner 2	galipette 3
galoper 3	gant 8	gants 1
garage 16	garage-la' 1	garc*on 37
garc*on-la' 1	garc*ons 8	garde 6
garder 3	gare 10	gare' 2
gare'es 3	garer 3	gauche 10
gaz 1	ge'nait 1	ge'nant 1
ge'ne 1	ge'rard 12	genou 3
genoux 8	genre 1	gens 25

gentil 17
gentille 13
gentils 1
gentiment 1
gilet 2
ginette 63
girafe 15
glace 3
glace'e 1
glaces 3
glisse 4
glisse' 3
glisser 4
gne' 1
gneu 2
gomme 9
gomment 1
gorge 3
gou't 2
goulue 1
gourmand 4
gourmande 1
goutte 4
gouttes 1
grain 1
grains 1
grand 31
grand-me're 2
grande 21
grandes 2
grands 3
gratte 1
gre've 5
grenouille 2
grenouilles 1
griffes 1
grille 1
grimpe 1
grimper 2
gris 10
grog 1
gronde' 1
gronomme 11
gros 89
groseilles 6
grosse 52
grosses 9
grossi 1
grue 8
gruye're 1
gue'ri 1
gue'rie 1
gueule 2
guidons 2
guidons-la' 1
guiges 2
guilis 1
guitare 9
guy 1
gyroscope 1
h 6
habille 3
habille' 3
habiller 1
habite 3
habitent 1
hache 1
halle 7
hamster 3
haut 17
he' 13
he'lice 10
he'licopte're 30
he'licopte'res 1
he'risson 5
hein 16
helena 1
herbe 7
heure 28
heures 6
hexagone 4
hi 5
hibous 1
hier 4
hippocampe 1
histoire 8
ho'tel 2
hop 34
hoquet 4
houp 1
house 1
hue 24
hui 32
huit 4
hum 3
huma 1
hyppolite 1
i 12
iaa 1
ici 21
ide'e 8
idiot 9
idiots 1
il 1012
ils 175
image 2
importe 4
index 1
indien 2
installe 2
installer 2
ion 1
ira 1
irait 1
iras 1
isabelle 3
j 358
jamais 2
jambe 2
jambes 9
janou 6
janoucha 1
japonaise 1
jardin 15
jaune 25
jaunes 8
je 1164
jean-michel 2
jean-paul 1
jeanine 3
jet 1
jete' 1
jeter 9
jetons 1
jette 3
jettent 1
jetterai 1
jeu 6
jeu-la' 2
jeudi 2
jeuh 2
joli 10
jolie 6
josiane 15
jouaient 1
jouais 1

joue 31
joue' 11
jouent 1
jouer 31
jouera 1
jouerai 1
jouerait 3
jouet 16
jouets 31
jour 11
journal 8
journe'e 1
judo 3
jumelles 4
jure 6
jus 6
jusqu 2
juste 1
kangourou 6
kira 1
klaxon 2
knickers 1
kola 1
l 619
la 1457
la' 871
la'-bas 24
la'-dedans 24
la'-dessus 9
la'-haut 19
la'r 1
labo 1
lacets 3
laid 1
laids 1
laine 5
laisse 4
laisse' 9
laisse'e 1
laisse'es 1
laisse-moi 2
laisser 1
laisses 1
lait 14
lala 1
lampe 5
lance 1
lance'e 2
lancer 9
langogne 12
langue 10
lapin 20
lapins 4
laquelle 3
lavabo 4
lave 5
lave' 1
laver 5
le 1641
le' 3
le'cher 1
le'go 4
le'gos 6
le've 3
le've-toi 1
lequel 9
les 777
lettre 1
lettres 1
leu 4
leuleu 2
leur 11
leurs 11
leve' 2
leve'e 5
lever 1
li 1
lieu 2
lilalalala 1
lili 1
liliane 1
limoges 7
lion 6
lions 4
lire 9
lis 4
lise 2
lit 27
livre 49
livres 10
locomotive 11
loin 8
loine 1
lolo 1
long 9
longs 1
longtemps 10
longue 2
losange 7
loup 38
loups 6
lourd 8
lourde 3
lourds 2
loze're 1
luge 5
luges 1
lui 40
lumie're 18
lumie'res 7
lune 17
lunette 1
lunettes 10
lunettes-la' 1
luxembourg 2
m 76
ma 68
machin 2
machine 33
machine-la' 2
machines 4
madame 12
madeleine 204
magasin 5
magasin-la' 1
magne'ophone 4
magne'tophone 14
magnifique 2
mai's 2
mai'tresse 22
mai'tresses 1
main 57
mains 16
maintenant 54
mais 106
maison 111
maison-la' 3
maisons 24
majeur 1
mal 60
malade 11
malades 3
malheur 4
malo 4
mama 1
mamaman 1
maman 204
mamie 89

man 1
mange' 37
mangeaient 2
manger 82
mani 1
manquait 1
manquer 1
mapa 2
marche 57
marcher 14
margoulette 2
marine 6
marque' 3
marron 12
marteau 11
masque 4
matin 4
me'canique 4
me'dicament 13
me'la 2
me'me're 3
me'tal 6
menton 2
merde 6
messieurs 2
mets-le 4
mette 1
mettrai 1
meubles 3
miam 12
michel 11
microbes 3
mienne 7
miettes 10
migrec 3
mimimorire 1
minouche 2
mise 1
mitalie 1
mobilis 1
modeur 1
mollets 1
monde 10
monsieur-la' 6
montagnes 1
montbrun 4
monte's 1
montes 1
montre-la 1
montrerai 2
mane'ge 20
mange'-cale 1
mangeait 1
mangeraient 1
manifestation 2
manque 8
manteau 7
marc 38
marche' 2
marco 8
mari 1
marionnettes 2
marraine 1
marrons 1
martine 6
masque-la' 1
mauvais 8
me'chant 11
me'dicaments 8
me'me 46
me'mes 2
me'tro 14
mer 9
mes 6
met 68
mets-les 1
mettes 1
mettre 178
meueu 1
miaou 3
mickey 4
mie' 3
miens 1
mieux 6
milieu 2
minet 1
mireille 3
mise'ricordes 3
mithalie 2
moche 1
moi 233
moment 1
monse 1
monsieurs 7
montaient 1
monte 39
montent 1
montparnasse 4
montre-moi 4
montres 5
mange 36
mange-cale 1
mangent 2
manges 3
manou 1
manque' 2
manteau-la' 1
marchand 2
marchent 1
marcominouche 1
marianne 1
marmotte 1
marrant 2
marseille 1
masou 3
mathalie 2
me 91
me'chants 1
me'got 1
me'me'me'seu 1
me'seu 1
menteur 3
merci 29
mesdames 1
mets 68
mettant 1
mettra 1
meuble 1
meuh 2
michambe 2
micro 6
mien 12
miette 5
migre 2
mimi 1
minou 5
mis 66
misette 40
mobi 1
moches 1
moitie' 6
mon 114
monsieur 143
montagne 16
montbrin 1
monte' 11
monter 23
montre 61
montrer 13
montsouris 2

monument 1	moque 2	morceau 22
morceau-la' 1	morceaux 6	morde 3
mordes 3	mordu 2	mort 3
morte 1	morts 3	mot 1
moteur 26	moteur-la' 1	moteurs 7
moto 85	moto-la' 2	motos 4
mots 2	mouche 2	mouche' 1
mouchelette 3	mouchelettes 1	mouches 3
mouchoir 14	moudre 4	mouds 1
mouflon 1	mouille' 3	mouille'e 6
mouille's 1	mouiller 1	moules 1
moulin 4	moulineaux 2	moulins 1
moulu 1	moustache 1	moustaches 1
mouton 14	moutons 10	moyen 4
mum 4	mur 4	mur-la' 1
murs 1	muse' 1	muse'se'se' 1
muse'se'se'seau 1	muse'se'seau 1	muse'seau 1
museau 1	musicien 1	musique 56
musiques 4	myriam 10	myrtilles 4
n 11	na 1	nage 2
nageoires 1	nager 3	nallumettes 1
nanimals 1	nanou 2	narbre 2
nautres 2	navion 13	navions 1
ne 17	neige 26	nen 11
nenfant 4	nettoie 4	nettoye' 1
nettoyer 3	neuf 3	neus 1
neuse 1	nez 12	ni 1
nicole 4	nid 1	no'tre 2
noe'l 1	noie 2	noir 47
noiraud 3	noire 4	noirs 4
noiseau 1	noix 1	nom 23
non 1137	normal 1	nours 2
nous 3	nouveau 2	nouvelle 2
noye' 1	nu 4	nuage 2
nue 2	nuit 13	nulle 1
nume'ros 1	nus 4	
o 1	oblige's 1	obstine 1
ocarine 2	occupe 1	oeil 2
oeils 1	oeuf 6	oeufs 5
offre 2	oh 326	oignons 1
oiseau 7	oiseaux 8	ombre 3
omm 1	on 520	ont 41
ope're' 1	orange 14	oreille 6
oreiller 1	oreilles 7	orsay 8
otite 1	otites 2	ou 24
ou' 260	ouah 7	ouais 18
oubliais 2	oublie 1	oublie' 7
oublie'e 1	ouf 1	ouh 7
oui 1351	oui'e 6	ouie' 1
ours 10	outils 7	ouvert 9

ouverte 5	ouvertes 7	ouverts 1
ouvrai 1	ouvre 20	ouvre-moi 2
ouvres 5	ouvrir 24	ovale 2
pa 3	pa'te 1	paf 1
paff 1	page 6	pages 2
pain 15	pam 2	pan 28
panier 7	panne 4	panneau 5
panneaux 1	panpan 1	pansement 1
pantalon 8	papa 277	papas 1
papier 27	papiers 6	papiers-la' 1
papillon 8	papillons 4	paquet 4
paquets 2	par 84	par-dessus 3
parai't 1	parapluie 2	parc 3
parce 110	pardon 6	pareil 6
parents 3	parfait 1	parfois 5
paris 1	parking 3	parle 11
parler 20	parles 1	pars 3
part 5	partage 1	partent 3
parti 15	partie 19	parties 2
partir 12	partira 1	partiras 2
partis 1	partout 6	pas 1060
pascal 10	passage 4	passe 21
passe' 3	passe'e 2	passer 7
pastroui'llarare 1	pastroui'llare' 1	pastrouillai'ra 1
pastrouillai're' 1	pastrouilles 1	pataboum 1
patin 3	patins 2	patrick 3
patte 2	pattes 5	pauvre 8
pauvres 5	pe 1	pe'che 1
pe'cheux 1	pe'dale 1	pe'se 2
pe'tales 3	peau 22	peigne 15
peigner 6	peignes 1	peine 1
peinture 3	pelle 14	penche 1
pendant 1	pense' 2	penses 1
penses-tu 4	perd 2	perdre 3
perdu 25	perdue 1	perroquet 1
perroquets 1	personne 17	personnes 1
petit 308	petit-de'jeuner 1	petite 189
petites 38	petits 90	peu 67
peuh 4	peur 13	peut 45
peut-e'tre 14	peuvent 2	peux 33
phare 3	phares 1	philippe 105
philou 4	photo 9	photos 9
piano 7	pie'ce 3	pie'tons 1
pied 25	pieds 15	pierre 1
pierres 2	pierrot 1	pigeon 1
pin 1	pinceau 3	pingouin 1
pinocchio 1	pipe 9	pipi 27
piqu're 3	piquants 2	pique 7
pique' 1	piquent 1	piquer 6
pis 3	piste 3	pistolet 19

pistolets 4
plaf 1
plai't 24
plancher 6
plas 1
plateau 3
pleine 2
pleurer 2
plier 3
pluie 5
plusieurs 6
poche 9
pof 2
pointe 6
poire 2
poissons 13
pomme 6
pompiers 4
porte 52
porte-monnaie 6
portes 13
pose-les 1
possible 5
poubelle 10
pouf 2
poulet 1
poupe'e 8
pour 445
pourrais 4
pousse 5
poussie'res 4
pouze 2
pre'pare 2
pre'te-moi 1
prend 10
prendra 3
presque 11
prochaine 1
promener 8
publicite' 1
puisque 4
puzzle 9
qu 444
que 233
quelles 1
quels 1
quille 2
raccourcie 2
raconter 1
radio 1

placard 1
plafond 3
plaisantin 1
planques 2
plastique 3
plates 2
pleure 3
pleut 3
plof 1
plume 2
pneu 7
poche-la' 1
poi 6
points 2
pois 1
police 7
pomp 1
pont 14
porte' 6
portent 2
portie're 1
poser 3
postale 2
poubelles 1
poule 13
poum 4
poupe'es 3
pourquoi 169
pourrait 2
pousser 2
poussin 2
pouzle 1
pre's 1
premie're 3
prendait 2
prendre 28
pris 21
profondeur 1
prote'ger 1
puis 55
pull 1
puzzle-la' 4
quand 60
quel 14
quelqu 6
queue 23
quilles 2
racheter 6
racontes 1
radotes 1

place 16
plage 3
planche 1
plantes 8
plat 2
plein 44
pleure' 7
plie 1
plouf 4
plus 159
pneus 3
pochette 1
poids 2
pointu 2
poisson 17
poltron 3
pompier 7
porquerolles 3
porte-moi 1
porter 8
pose 3
positoire 1
pot 14
pouce 3
poules 4
poumandarinette 13
poupou 10
pourra 1
pourront 1
poussie're 7
poussins 2
pre'au 1
pre't 15
premier 3
prende 1
prends 20
proble'mes 4
prome'ne 6
pu 4
puisqu 1
pull-over 2
puzzles-la' 3
quatre 8
quelle 28
quelque 28
qui 348
quoi 334
raconte 1
radiateur 2
rails 19

raisin 2	raisins 1	raison 1
ramasse 12	ramasse' 3	ramasse'e 1
ramasser 8	ramasses 3	rames 1
range 6	range' 3	ranger 10
rappelle 1	rapporte' 1	raquette 25
ras 1	raser 2	rat 1
rate 3	rateau 12	rateaux 1
rati 2	rats 2	rattrape 1
rattrape' 1	re' 1	re'chauffer 1
re'fle'chis 2	re'fle'chit 1	re'galer 2
re'pare 5	re'pare' 8	re'pare'e 1
re'parer 11	re'ussi 1	re'veil 4
re'veille 1	realle's 2	rec*u 1
recacheras 1	recasse' 4	recasser 3
recoller 1	reconnai't 5	reconnais 1
reconnu 3	reconstruit 1	recoucher 1
rectangulaire 1	reculotte 1	rede'monter 1
redorme 2	reenle've 2	refaire 6
refais 6	refait 1	refermer 1
refermes 1	refroidi 1	refroidisse 2
regarde 209	regarde' 5	regardent 4
regarder 12	regardes 1	regardez 5
regardons 1	relaver 1	releve'e 1
remet 3	remets 6	remets-la 1
remettez-les 1	remettre 8	remis 2
remmene' 1	remonte 3	remonter 2
remorque 9	remplacer 1	remue 2
renard 10	rends 2	rentre 13
rentre' 5	rentre'e 4	rentrer 12
rentrera 1	rentrerait 1	renverse 3
renverse' 4	renverse'es 1	renversent 1
renverser 2	renvoyer 1	repartie 1
repartir 3	ressemblent 3	ressort 3
ressortir 1	ressorts 1	restaurant 1
reste 8	reste' 2	restent 1
rester 7	resteras 1	restes 7
retombe 3	retombe' 2	retourer 2
retourne 5	retourner 5	retrouve 1
retrouve' 6	retrouver 1	retrouvera 1
revenue 1	reviendra 2	reviens 2
revient 5	revoila' 1	rhabiller 4
rhabillette 3	rhinoce'ros 2	rien 42
rigolo 5	rire 1	risquais 1
rit 4	rivie're 2	riz 1
ro 2	robe 5	robert 1
robes 1	robinet 1	robinets 1
robot 4	roche 1	rocher 4
rochers 2	rogalski 8	rond 13
ronde 5	rondelles 1	rondes 2
ronds 8	roros 1	rose 8

roses 7
rou 10
roue 37
roue-la' 3
roues 51
roues-la' 2
rouge 73
rouges 13
roule 53
roule' 7
rouleau 3
roulent 4
rouler 26
roulettes 2
roulez 3
roum 9
rououououou 2
route 21
routes 3
rue 19
s 174
sa 63
sable 12
sac 37
sage 11
saint-cloud 1
sais 192
sait 6
sale 3
sales 4
salete's 4
sali 1
salir 5
salit 4
salle 1
salon 8
salut 1
samedi 4
samedie' 2
samos 2
sandales 1
sang 1
sans 6
sante' 2
saperlipopette 1
sapin 1
sar 3
saute 7
saute' 3
sautent 1
sauter 1
sauve 1
sauver 2
savent 1
savoir 2
savon 2
scania 1
scie 4
sciure 1
scoch 4
se 97
se'che 4
se'chent 1
se'que' 2
se'que's 1
seau 5
second 1
secoue 2
seguin 6
seize 1
sel 1
semaine 2
semblant 3
sens 19
sens-la' 1
sent 5
sera 3
serai 1
serait 2
seront 1
serpent 8
serre 5
sert 10
servent 1
serviette 6
serviette-la' 1
serviettes 1
servir 2
ses 29
seul 49
seule 19
seulement 5
shampooing 1
si 137
sie'ge 14
sieste 2
sifflet 1
singe 13
singes 1
sinon 1
sire'ne 1
sirop 1
six 2
ski 2
skis 4
so 1
soif 2
soigne 4
soigne' 3
soigner 1
soignera 1
soignerait 1
soir 1
soleil 28
solide 3
sommeil 2
son 67
sonne' 1
sont 117
sophie 1
sors 2
sort 15
sortant 1
sorte 2
sorti 7
sortie 2
sortir 15
sortira 2
sortirai 1
sou 2
souffe 1
souffle 2
souffler 5
souffleu 3
soule've 1
souleve' 1
soupe 3
sourcils 5
sourires 1
souris 36
sous 30
souvent 1
souviens 10
soye 6
soyent 1
spirou 1
statue 1
stylo 13
stylos 3

su 1	sucre 49	sucre' 1
sucres 9	suffit 2	suis 54
suite 4	super 1	suppositoire 1
suppositoires 1	sur 307	surprise 1
surtout 2	symbolique 1	syndicat 4
t 152	ta 49	tabe' 5
table 54	tableau 12	tableaux 4
tabouret 6	tache 1	tache' 3
tache'e 2	taille 4	taille-crayon 2
tailler 20	taintain 1	talle' 5
tambour 11	tant 2	tape 4
tape' 3	tapis 5	tartine 12
tas 10	tase 4	tasse 2
tata 2	taureau 9	taxi 2
te 83	te'gon 1	te'le'phe'rique 4
te'le'phone 12	te'le'phone' 1	te'le'phoner 2
te'le'vision 9	te'le'visions 2	te'te 27
te'tes 3	teau 2	temps 6
temps-la' 1	tenait 1	tenir 3
tente-la' 1	ter 2	terre 50
tes 9	the' 1	tic-tac 1
ticket 1	tickets 1	tien 4
tienne 4	tiennent 2	tiens 56
tient 20	tige 8	tigre 7
timbale 5	timbales 1	tinette 1
tire 17	tire' 2	tire-bouchon 2
tirelire 2	tirer 6	tires 1
tiroir 13	tiroirs 3	tissus 2
to't 1	toboggan 36	toboggans 4
toffe 2	toi 74	toilette 3
toit 16	toits 1	tom 1
tombe 30	tombe' 72	tombe'e 20
tombe'es 1	tombe's 1	tombent 5
tomber 40	tomberas 1	ton 85
tonne 3	tonneau 3	tonneaux 2
tonnerre 7	tord 5	tordu 1
tordue 2	tordues 1	tortille 3
tortue 18	tortues 1	touche 9
touche' 3	toucher 4	touf 1
touille 1	toujours 8	toupie 1
tour 16	tour-la' 1	tourne 41
tourne-vis 5	tournent 2	tourner 22
tourterelle 3	tourterelles 3	tous 51
tousse 6	tousser 5	tousses 1
tout 199	toute 25	toutes 27
toux 2	traces 8	tracteur 34
tracteurs 4	trai'ne' 2	trai'ner 1
train 110	train-la' 1	trains 11
traire 2	tramway 2	tranquille 1
transforme' 1	transparente 1	transporte 3

travail 8
travaille 6
travaille' 3
travaillent 3
travailler 17
travailles 3
travers 9
tre's 50
treize 1
trempe 1
tremper 2
trempes 1
triangle 25
triangles 3
triches 2
triste 2
trois 42
troisie'me 1
trombone 1
trompe 2
trompe' 2
trompette 3
tronc 1
trop 52
trotinette 1
trotinettes 1
trottoir 7
trou 38
troue's 1
troup 3
trous 16
trous-la' 1
trouve 6
trouve' 9
trouve'e 1
trouve's 1
trouver 1
truc 18
truc-la' 3
trucs 15
trucs-la' 2
tu 374
tube 1
tue 2
tuer 4
tunnel 1
tuser 1
tut 11
tuyau 5
tuyaux 1
tyran 1
u 1
un 1079
une 771
use' 5
usine 1
va 287
va-t 25
vacances 2
vache 5
vaches 15
vague 1
vais 262
valise 5
valises 5
vas 47
ve' 1
ve'lo 18
ve'los 7
ve'tement 5
ve'tements 4
veau 2
veaux 3
vend 2
venir 6
vent 13
ventilateur 8
ventre 16
venu 4
venue 1
venus 1
ver 2
verrai 1
verras 1
verre 25
verres 5
vers 1
verserai 1
vert 36
vert-la' 1
verte 8
vertes 1
verts 3
verture 1
veste 1
veulent 6
veut 52
veux 311
veux-tu 3
vide 4
vider 4
vie 2
vieille 4
vieilles 2
viendra 1
viendras 1
viendront 2
vienne 6
viennent 1
viens 30
vient 19
vieux 10
vilain 4
vilaine 6
village 1
ville 2
vin 6
vins 7
violet 8
violettes 1
violon 1
vis 9
visse 1
visse-le 1
visser 1
vit 1
vite 14
vitre 17
vitres 6
voi 2
voila' 174
voile 3
voir 102
vois 88
voisin 2
voit 16
voiture 270
voiture-la' 8
voitures 85
voitures-la' 1
volant 13
vole 10
vole' 1
volent 3
voler 3
vomi 1
vomir 1
vont 9
voudrais 4

voudrait 1	voulais 2	voulez 4
voulu 2	vous 13	voyons 1
vrai 19	vraie 5	vraies 4
vraiment 3	vrais 2	vroum 2
vu 43	vue 3	
wagon 12	wagon-la' 1	wagons 5
xagone 1		
y 345	yaourt 3	yeux 31
zailes 2	zarrivent 1	ze'coutent 1
zencore 1	zigzag 1	zizi 3
zoiseau 2	zont 1	zoo 4
zoos 3	zorro 5	zoutils 3
zut 7		
&&& 1370		
0h 1		
10 1		

Types = 2698
Tokens = 56,982

APPENDIX D

Dictionary I

Table 38 gives, for each lexical category, the list of words that were classed in that category in Dictionary I. Table 39 gives a rank summary of the categories, and Table 40 gives an alphabetical summary of the categories.

TABLE 38

Lexical Categories of Dictionary I

classification = nc1.0   number of words = 526

a'ne abat-jour accident accidents accorde'on aigle air
alligator amour anneaux annulaire ape'ritif appareil
appartement arbre arbres arc-en-ciel arcs-en-ciel argent
autobus autocar autocars avion avis ba'ton bain baiser
baisers balai balais balcon ballon ballons banc bancs
barreau barreaux bassin bateau bateaux be'be' be'be's
be'ton bec besoin biberon bigoudis billet bitonniau ble'
bobo bol bols bonbons bonhomme bord bouchon bouton boutons
bouts bracelet bracelets bras briquet bruit bureau cabinet
caca cache-cache cadeau cadeaux cafe' cahier cailloux
calembour camion camions camping canard canards carnet
carreau carrelage cartable casque castor cendrier cerf
cha'teau chameau chamois champignon champignons champs
chapeau chapeaux chaperon charriot charriots chasseur chat
chats chauffe-eau chausson chaussons cheveu cheveux
chichis chien chiens chiffon chigne chocolat chou ciel
cigares cine'ma cirage ciseaux clown clowns co'te's codile
coffre coin collier coloriages commerce conduiteur copains
coquillages cornichon cou coude coudes couloir coup coups
coussin coussins couteau couteaux couvercle crayon crayons
crochet crochets crocodile croissant croissants cube cygne
de' de'but demi-rond demi-soleil dentifrice dessin dessins
dimanche disque disques docteur dodo doigt doigts dos
dossier drabon dragon drap drapeau drapeaux e'clair
e'couteur e'crou e'cureuil e'dicaments e'le'phant
e'le'phants e'lectrophone e'pouvantail e'tage e'tages

```
e'te' e'tudiants edelweiss ele'phant embouteillage endroit
ennuis escalier escaliers facteur fauteuil fauteuils feu
feux fil filet film fils flic foie fourneau fourreau
fre're frigidaire fromage front fruits funiculaire fusil
fusils ga'teau ga'teaux gag gages gant gants garage
garc*on garc*ons gaz genou genoux genre gilet gou't grain
grains grog guidons guilis gyroscope hamster he'licopte're
he'licopte'res he'risson hexagone hibous hippocampe ho'tel
hoquet index jardin jet jetons jeu jeudi jouet jouets jour
judo jus kangourou klaxon knickers kola labo lacets lait
lapin lapins lavabo le'go le'gos lieu lion lions lolo
losange loup loups machin magasin magne'ophone
magne'tophone mai's malheur mane'ge manteau marchand mari
marrons marteau masou masque matin me'dicament
me'dicaments me'got me'tal me'tro menton meuble meubles
micro microbes minet minou modeur mollets moment monde
monsieur monument morceau morceaux mot moteur moteurs mots
mouchoir mouflon moulin moulins mouton moutons mur murs
museau narbre navion navions neus nez nid noe'l noiseau
nom nours nuage nume'ros ocarine oeil oignons oiseau
oiseaux oreiller ours outils pain panier panneau panneaux
pansement pantalon papier papiers papillon papillons
paquet paquets parapluie parc parents parking passage
patin patins pe'tales perroquet perroquets petit-de'jeuner
phare phares piano pie'tons pied pieds pigeon pin pinceau
pingouin pipi pistolet pistolets placard plafond plancher
plateau pneu pneus poids points pois poisson poissons
pompier pompiers pont porte-monnaie positoire pot pouce
poulet poussin poussins pre'au proble'mes pull pull-over
puzzle radiateur rails raisin raisins rat rateau rateaux
rats re'veil renard ressorts restaurant rhinoce'ros riz
robinet robinets robot rocher rochers rouleau sable sac
salon samedi samedie' sang sapin savon seau sel serpent
shampooing sie'ge sifflet sirop ski skis so soir soleil
sommeil sou sourcils sourires stylo stylos suppositoire
suppositoires syndicat tableau tableaux tabouret
taille-crayon tambour tapis tas taureau taxi
te'le'phe'rique temps the' ticket tickets tigre
tire-bouchon tiroir tiroirs tissus toboggan toboggans toit
toits tonneau tonneaux tonnerre tourne-vis tracteur
tracteurs train trains tramway triangle triangles trombone
tronc trottoir trou trous truc trucs tube tunnel tuyau
tuyaux tyran ve'lo ve'los ve'tement ve'tements veau veaux
vent ventilateur ventre ver verre verres village vin vins
violon voisin wagon wagons xagone yaourt zigzag zizi
zoiseau zoo zoos zoutils


classification = nc2.0   number of words = 381
```

ae'rogare agrafe aiguille aiguilles ailes allumette allumettes ambulances ancre antenne antennes araigne'e armoire assiette assiettes autoroute autos bagnole bague baguette baguitte balanc*oire balle balles banguette barrie're barrie'res base be'tise be'tises be'toleu be'toleuse be'tonneuse beguitte benne biche biches bille billes biscotte biscottes blanquette bobine boi'te boi'tes bombe bosse bottes bouche boucherie boue'e bouette bougie bougies boule boules bousse bouteille bouteilles boutiques branches brioche brouette bulle bulles cabane cabines cabriole cacahue'te cacahue'tes cage cages caisse campagne campagnes canne caravane caravanes carte cartes casquette casquettes casserole casseroles cendre cerise cerises cervelle chai'ne chaise chaises chambre chandelle chanson chansons chaussettes chaussures che'vre che'vres chemine'e chemine'es chemise chose choses cigarette cigarettes circulation classe clef clefs cloche colique commissions compote confiture confitures construction coqueluche corbeille corde cordes corne couleur couleurs cour course couture couverture craie cravate cravates cre'me cre'pe cre'pes crou'te cuille're cuille'res cuiller cuisine cuisse cuisses culotte dame dames dent dents e'chelle e'cole e'glique e'glise e'pe'e e'ponge e'tamines e'tiquette e'toiles eau encre espe'ce estatues fac*on face facture faim fanfare farine faucille faute fe'te fene'tre fene'tres fesse'e fesse'es fesses festation feuille feuilles ficelle fifille fille filles fin flamme fle'che fleur fleurs flu'tes fois fourchette fourrure fronde fuse'e galipette gare girafe glace glaces gomme gorge goutte gouttes grand-me're gre've grenouille grenouilles groseilles grue gueule guitare hache halle he'lice herbe heure heures histoire ide'e image jambe jambes journe'e jumelles laine lampe langue lettre lettres locomotive luge luges lumie're lumie'res lune lunette lunettes machine machines mai'tresse mai'tresses main mains maison maisons manifestation margoulette marionnettes marmotte marraine mer miette miettes mise'ricordes montagne montagnes moto motos moustache moustaches musique musiques myrtilles nageoires nallumettes noix nuit ombre oreille oreilles otite otites pa'te panne patte pattes pe'dale peau peine peinture pelle personnes peur photo photos pie'ce pierres pipe piqu're piste plage planche pluie poche pochette pointe poire police pomme portie're poubelle poubelles poule poules poupe'e poupe'es poussie're poussie'res profondeur publicite' queue quille quilles radio raison raquette remorque rivie're robe robes roche rondelles roue roues roulettes route routes rue salete's salle sandales sante' scie sciure semaine serviette serviettes sieste sire'ne soif soupe statue surprise table tache tartine tasse te'le'vision te'le'visions te'te te'tes terre tige

```
timbale timbales tirelire toffe toilette tortue tortues
toupie tourterelle tourterelles toux traces trompette
trotinette trotinettes usine vache vaches valise valises
verture veste vie ville vis vitre vitres voile voiture
voitures zailes


classification = vt0   number of words = 169

accompagne'e accroche' accroche'e accrocher achete'
achete'e achete'es achete's acheter agrafe's agrafer
allume' allume'e allumer amener amuser appele' apporte'
apporte's apporter apprendre arranger arroser asseoir
attacher attaque'e attaquer attendre attrape' attraper
baisse' baisse'e baisser balaye'es batter battre boire
brosser ca'liner cacher chercher coince' colorier conduire
construire coudre couper crotte' de'barbouiller de'bouche'
de'chirer de'croche' de'faire de'fendre de'molir de'monte'
de'monte'e de'monter de'shabille' demander dessiner dire
e'coute' e'couter e'craser e'crire e'teindre emmener
emporte' emporte's emporter enferme' enlever envoyer
essayer essuyer expliquer fabrique' fabriquer fascine'
foutre froisse' frotter gagne' gagner garder gare' gare'es
garer habiller installer jete' jeter lancer laver le'cher
lire mange' manger mettre montrer mouche' moudre mouiller
nettoye' nettoyer oublie' oublie'e pastroui'llarare
pastroui'llare' pastrouillai're' peigner perdre porte'
porter promener prote'ger racheter raconter ramasser
ranger rapporte' raser rattrape' re'chauffer re'galer
re'pare' re'pare'e re'parer recoller rede'monter refaire
refermer regarde' regarder regardez relaver remettre
remmene' remplacer renverser renvoyer retrouve' retrouver
rhabiller salir sauver savoir servir soigne' soigner
souleve' tailler te'le'phone' te'le'phoner touche' toucher
traire transforme' trompe' trouve' trouve'e trouve's
trouver tuer tuser vider visser vomir


classification = vt1   number of words = 158

accroche accrochent accroches ache'te aides aime allume
ame'ne amuse amusent appelle appellent appelles apprend
asseoit assieds attend attends attrape avale baisse balaie
bat boit bouscule cace cache caches chatouille cherche
cherches coince coincent comprends conduit connais
construit coud coupent couvre de'barbouille de'chire
de'fais de'moli de'montes de'passe de'passent de'pe'che
de'shabille de'visse demande dessine dessines devine
dirige dis disputent dit e'clabousse e'coute e'coutes
e'cris e'teins e'teint e'tonne embe'te embe'tes emme'ne
```

emporte endort enferme enle've enle'ves ennuie enregistre
entend entends envoie essaie essaye essuie explique
gomment habille installe jette jettent jure lave lis mange
mange'-cale mange-cale mangent manges met mets mouds
nettoie noie occupe oublie pastrouilles perd plai't
planques portent pre'pare prome'ne raccourcie raconte
racontes ramasse ramasses range rappelle rate rattrape
re'pare re'veille reconnai't reconnais reconstruit
reculotte reenle've refais refait refermes regarde
regardent regardes remet remets rends renversent
ressemblent retrouve sais sait sali salit sauve secoue
sent sert soigne soule've suffit tord touche touille
transporte trouve tue vend visse vomi ze'coutent

classification = ms   number of words = 119

abricure ajourabe ani ba bab babi baboum bate bates be'leu
beleu beu bibibibibobodidabe bibibobabibi bibiboboline
bibibobolune boiboiboiboi bolle' bolo bou boudie bouloum
boulu brou burode c* cass ce' ce'ce's che chonchon
chonchons coge' cor cre'p crip cro croi croisse crole da
dada de'de' droi e e' e'c escalou escoulou fan flip
foievelle ga gne' gneu guiges h house hui i ion la'r le'
leu li malo man mani mapa me'la me'me'me'seu me'seu
michambe mie' migrec mimi mimimorire mobi mobilis monse
muse' muse'se'se' muse'se'se'seau muse'se'seau muse'seau
na neuse o omm pa pe poi pomp poumandarinette poupou pouze
pouzle re' rhabillette ro roros rou sar scoch souffleu
tabe' taintain tase tata te'gon teau ter tinette tonne
troup u ve' voi _

classification = vn0   number of words = 89

appuyer arre'te' arre'te'e arre'ter arre'tez avancer
baigne' baigner balancer bru'ler casser chante' chanter
chasse' chasser chauffer cogne' cogner colle' colle'e
coller compte's coucher coule' couler courir crier croque'
croquer cuire de'coller de'gringole' descendre donne'
donne'es donner e'clater fermer finir freine' fumer
glisse' glisser grimper gronde' joue' jouer lever manquer
monte' monte's monter ouvrir parler passer pense' piquer
pleure' pleurer plier poser pousser prendre recasser
recoucher remonter rentre' rentrer retourer retourner
rouler roulez saute' sauter sonne' sortir souffler tape'
tenir tire' tirer tourner trai'ne' trai'ner travaille'
travailler tremper voir voler

classification = vn1   number of words = 85

appuie appuies arre'te arre'tent arre'tes bouge bru'le
bru'lent casse chante chantent compte couche coule crois
croit dansent descend descends discutent donne donnes
fauche ferment freine fume fumes glisse gratte grimpe
grossi habite habitent jouent le've manque monte montent
montes ouvre ouvres parle parles passe pe'se penche penses
pique piquent pleure plie pousse prend prends re'fle'chis
re'fle'chit refroidi remonte remue rentre retourne roule
roulent saute sautent se'chent sors sort souffe souffle
tape tient tire tires tortille tourne tournent travaille
travaillent travailles vit vois voit vole volent

classification = ko   number of words = 75

ah ai'e aie'e aiou allo aouiii babababababa bah be' be'e'
ben beuh blablablabla boh bouboubou bouh boum boup broum
clic crocrocro croooh croq dieu dis-donc domdomdomdom
e'e'h eh euh he' hein hi hop houp hue hum iaa jeuh lala
leuleu lilalalala merde meueu meuh miam miaou mum oh ouah
ouf ouh oui'e ouie' paf paff pam pan panpan pataboum peuh
plaf plof plouf pof pouf poum roum rououououou
saperlipopette tic-tac touf tut vroum zut Oh

classification = np2   number of words = 40

agne's allemagne angleterre be'ne'dicte carole clara
danielle ginette helena huma isabelle janou janoucha
jeanine josiane lili liliane loze're madeleine mama
mamaman maman mamie manou marianne marseille martine
mathalie me'me're mireille misette mitalie mithalie
mouchelette mouchelettes myriam nanou nicole scania sophie

classification = nc1.0 ad   number of words = 40

ballon-la' bateau-la' bol-la' bout-la' bouton-la'
boutons-la' bouts-la' bruit-la' camion-la' chapeau-la'
chien-la' co'te'-la' collier-la' crayon-la' creux-la'
dessin-la' e'tage-la' feu-la' ga'teau-la' garage-la'
garc*on-la' guidons-la' jeu-la' magasin-la' manteau-la'
masque-la' monsieur-la' morceau-la' moteur-la' mur-la'
papiers-la' puzzle-la' puzzles-la' sens-la' temps-la'
train-la' trous-la' truc-la' trucs-la' wagon-la'

classification = np1   number of words = 37

aristochats babar bambi bill bussy cognac franc*ois ge'rard gronomme gruye're guy hyppolite jean-michel jean-paul luxembourg marc marco marcominouche michel mickey minouche noiraud papa papas pascal patrick philippe philou pierrot pinocchio rati robert samos seguin spirou tom zorro

classification = aq0.0.2   number of words = 37

bizarre bleue bleues carre'e carre'es cloute' croise'e de'sole' difficile e'lectrique e'lectriques fragile glace'e goulue jaune jaunes juste marron me'canique moche moches mouille' mouille'e mouille's nu nue nus pointu possible rectangulaire rouge rouges sage sale sales symbolique tranquille

classification = aq0.0.2,vt0   number of words = 35

abi'me'e attache' blesse' cache' cache'e complique' coupe' coupe'e coupe's de'chire' de'coupe' de'tache'e dessine' dessine'e dore' efface' enleve' enleve'e enleve's envoye' habille' lave' noye' ope're' ramasse' ramasse'e range' releve'e renverse' renverse'es sucre' tache' tache'e troue's use'

classification = nc2.0 ad   number of words = 31

aiguille-la' auto-la' baguette-la' barres-la' be'te-la' be'tes-la' boi'te-la' bouteille-la' chambre-la' che'vre-la' chose-la' crou'te-la' cuille're-la' cuiller-la' dame-la' eau-la' famille-la' ficelle-la' fille-la' fleur-la' lunettes-la' machine-la' maison-la' moto-la' poche-la' roue-la' roues-la' serviette-la' tente-la' voiture-la' voitures-la'

classification = nc2.0,vt1   number of words = 31

attache bande bandes barbe barres brosse brosses caresses cire coupe coupes envie fabrique forme ge'ne griffes lance montre montres mouche mouches offre pe'che place plantes plume porte portes serre taille trompe

```
classification = aq1.0.2   number of words = 27

be'ta blanc blancs blonds brun compresseur content
contents crasseux de'faits embe'tant entier entiers
fatigant gourmand gris laid laids marrant pareil ras
se'que' se'que's vert verts vilain violet


classification = aq1.0.0   number of words = 26

ancien chaleureux chaud chauds dangereux doux dur gentil
gentils grand grands long longs lourd lourds mauvais
me'chant me'chants nouveau parfait petit petits plein
pre't seul vieux


classification = aq1.0.2,nc1.0   number of words = 25

blagueur carnassier carnassiers cochon cochons coquin
creux droit fou fous froid idiot idiots majeur menteur
morts moyen musicien piquants plaisantin plat poltron
rigolo rond ronds


classification = vi1   number of words = 23

arrive baille de'raille dors dort existe fabouilles
marchent migre pars radotes retombe reviens revient rit
tombe tombent tousse tousses triches viens vient zarrivent


classification = aq0.0.2,vn0   number of words = 22

bru'le' bru'le'e casse' casse'e casse'es casse's couche'
couche'e creve' fatigue' ferme' ferme'e ferme's leve'
leve'e manque' marque' passe'e pique' recasse' roule'
vole'


classification = np0   number of words = 21

assi-les-moulineaux boirat boulogne carqueranne demassieu
demeussieu eiffel kira langogne limoges montbrin montbrun
montparnasse montsouris moulineaux orsay paris pe'cheux
porquerolles rogalski saint-cloud


classification = aq2.0.0   number of words = 17
```

chaude dernie're dure gentille grande grandes longue lourde nouvelle petite petites pleine prochaine seule vieille vieilles vilaine

classification = dv0 number of words = 15

ainsi alors beaucoup encore ensemble expre's ici jamais peut-e'tre plus seulement surtout tre's trop zencore

classification = vi0 number of words = 15

arriver de'jeune' dormir existe' galoper marcher nager partir realle's repartir ressortir retombe' tomber tousser venir

classification = aq2.0.2 number of words = 15

blanche blanches bouillante contente froide gourmande morde mordes morte plates postale se'che transparente verte vertes

classification = nc1.0,vt1 number of words = 12

beurre bois lit partage peigne peignes sens singe singes sucre sucres te'le'phone

classification = ac,nc1.0 number of words = 12

cinq deux dix dix-neuf douze huit quatre seize six treize trois 10

classification = aq0.0.0 number of words = 12

dro'le dro'les magnifique pauvre pauvres solide super triste vrai vraie vraies vrais

classification = nc0.0 number of words = 11

ami amis enfant enfants livre livres moules nenfant page pages tour

```
classification = nc2.0,vn1   number of words = 11

balance chasse claque colle fatigue figure grille joue
pose trempe trempes


classification = vn6   number of words = 11

courait croyais donnait fermais jouaient jouais manquait
montaient ouvrai prendait tenait


classification = nc1.2   number of words = 10

animaux chevals chevaux gens messieurs monsieurs nanimals
oeils oeufs yeux


classification = vt6   number of words = 10

attendais attrapait cherchais connaissais essayait ge'nait
mangeaient mangeait oubliais risquais


classification = vt5   number of words = 10

attende attendes boivent disent e'crise lise mette mettes
savent servent


classification = dv1   number of words = 10

de'ja' demain hier longtemps parfois puis souvent to't
toujours vite


classification = vt6,va6   number of words = 8
avaient avais avait devais faisaient faisais faisait
voulais

classification = vt10   number of words = 8
cacheras de'molira diras mettra pastrouillai'ra recacheras
retrouvera soignera

classification = vn9   number of words = 8
casserait fuirait jouerai jouerait rentrerait sortirai
verrai verserai

classification = ep0   number of words = 8
chez dans depuis par pour revoila' sans sur
```

```
classification = vt1,va1   number of words = 7
ai doit fais peut peux veut veux

classification = vi0,va0   number of words = 7
alle' alle's allez e'tre reste' rester talle'

classification = nc1.0,lc0   number of words = 7
au-revoir bonjour dommage merci milieu pardon salut

classification = vt9,va9   number of words = 7
aurait ferai ferait pourrais pourrait voudrais voudrait

classification = aq0.0.2,nc1.0   number of words = 7
bleu bleus carre' e'lastique e'lastiques ovale plastique

classification = vt9   number of words = 7
chercherai dirait jetterai mangeraient mettrai montrerai
soignerait

classification = lc1   number of words = 6
abord au-dessus la'-bas la'-dedans la'-dessus la'-haut

classification = nc1.1   number of words = 6
animal cheval chevau journal oeuf travail

classification = vn10   number of words = 6
arre'tera jouera prendra rentrera sortira verras

classification = vt21 pe15   number of words = 6
attrape-le cache-le de'fais-le envoie-le mets-le visse-le

classification = vt0,va0   number of words = 6
avoir faire laisse' laisse'e laisse'es laisser

classification = nc2.0,vi1   number of words = 6
bave marche nage part rames souris

classification = aq0.0.2,nc2.0   number of words = 6
be'te bouche'e marine orange rose roses

classification = vt20   number of words = 6
bu mordu rec*u reconnu remis su

classification = vt5,va5   number of words = 6
doivent font ont peuvent veulent zont

classification = vi20   number of words = 6
dormi failli parti partis repartie revenue
```

```
classification = aq0.0.2,vt20   number of words = 6
moulu perdu perdue tordu tordue tordues

classification = vi10   number of words = 6
partira partiras reviendra tomberas viendra viendras

classification = aq1.0.2,np1   number of words = 5
anglais chinois indien noir noirs

classification = aq0.0.2,vi0   number of words = 5
arrive' tombe' tombe'e tombe'es tombe's

classification = vn20   number of words = 5
couru descendu descendue sorti vu

classification = vt21 pe3   number of words = 5
de'fais-moi dis-moi montre-moi porte-moi pre'te-moi

classification = vi1,va1   number of words = 5
devient parai't restent suis vais

classification = aq0.0.2,vn1   number of words = 5
fini finis gue'ri gue'rie re'ussi

classification = vi15,va15   number of words = 4
aillent alles soye soyent

classification = dv2   number of words = 4
ailleurs par-dessus partout pre's

classification = dv3   number of words = 4
assez peu presque tant

classification = vt15,va15   number of words = 4
aye ayent fasse fasses

classification = dv4   number of words = 4
be'tement debout dro'lement gentiment

classification = aq2.0.1   number of words = 4
belle belles grosse grosses

classification = pi8,ai   number of words = 4
chaque me'mes plusieurs quelque

classification = nc1.0,dv2,ep0   number of words = 4
derrie're dessous dessus devant

classification = vn5   number of words = 4
descende prende refroidisse tiennent
```

```
classification = vr1   number of words = 4
envole moque obstine souviens

classification = vi2,va2   number of words = 4
es est va vas

classification = vt20,va20   number of words = 4
eu eue pu voulu

classification = lc0   number of words = 4
importe parce pis semblant

classification = vi10,va10   number of words = 4
ira iras resteras sera

classification = dv6   number of words = 4
n ne non plas

classification = vi5   number of words = 4
partent redorme vienne viennent

classification = vn12   number of words = 3
appuyant courant sortant

classification = vt10,va10   number of words = 3
aura auras pourra

classification = aq0.0.1,pi8,ai   number of words = 3
autre autres nautres

classification = nc2.0,vn0   number of words = 3
cogne'e fume'e rentre'e

classification = cs,dv0   number of words = 3
combien comme comment

classification = nc1.0,vi0   number of words = 3
de'jeuner marche' rire

classification = vi6,va6   number of words = 3
e'taient e'tais e'tait

classification = cc   number of words = 3
et ni ou

classification = vi9,va9   number of words = 3
irait serai serait

classification = dv5   number of words = 3
ouais oui vraiment
```

```
classification = aq1.0.2,vn20   number of words = 3
ouvert ouverts pris

classification = cs   number of words = 3
puisqu puisque sinon

classification = aq2.0.2,nc2.0   number of words = 3
ronde rondes violettes

classification = nc1.0,vt12   number of words = 2
aimant e'clant

classification = nc2.0,vi0,va0   number of words = 2
alle'e alle'es

classification = aq1.0.0,vt12   number of words = 2
amusant e'tonnant

classification = vi11   number of words = 2
arriveront viendront

classification = aq1.0.2,vt20   number of words = 2
assis mis

classification = nc2.0,ko   number of words = 2
attention flu'te

classification = ep5   number of words = 2
au aux

classification = vt11,va11   number of words = 2
auront pourront

classification = aq1.0.1   number of words = 2
beau gros

classification = nc1.0,vi1   number of words = 2
bout ressort

classification = pd2   number of words = 2
celle celles

classification = pd7   number of words = 2
celle-la' celles-la'

classification = vt3   number of words = 2
cherchons regardons

classification = aq0.0.0,nc2.0   number of words = 2
chouette vague
```

```
classification = nc1.0,lc1   number of words = 2
co'te' envers

classification = aq1.0.2,vt1   number of words = 2
de'fait e'crit

classification = nc1.0,dv2,lc1   number of words = 2
dedans dehors

classification = vn21 pe3   number of words = 2
donne-moi ouvre-moi

classification = pe10   number of words = 2
elle elles

classification = pe19,dv2,ep0   number of words = 2
en nen

classification = nc0.0 ad   number of words = 2
enfants-la' tour-la'

classification = vt12   number of words = 2
enlevant mettant

classification = lc3   number of words = 2
est-c*a est-ce

classification = vt21 pe3,va21 pe3   number of words = 2
fais-moi laisse-moi

classification = aq2.0.2,vt4,va4   number of words = 2
faite faites

classification = vm1   number of words = 2
faut pleut

classification = pe9   number of words = 2
il ils

classification = pe1   number of words = 2
j je

classification = aq2.0.2,np2   number of words = 2
japonaise noire

classification = aq0.0.1   number of words = 2
joli jolie

classification = nc2.0,vt1,va1   number of words = 2
laisse laisses
```

```
classification = pe18,po9,ao14   number of words = 2
leur leurs

classification = dv2,lc0   number of words = 2
loin loine

classification = pe2   number of words = 2
m me

classification = aq0.0.2,nc0.0   number of words = 2
malade malades

classification = nc2.2   number of words = 2
mesdames vacances

classification = po1   number of words = 2
mien miens

classification = vt21 pe16   number of words = 2
montre-la remets-la

classification = aq2.0.2,vn20   number of words = 2
ouverte ouvertes

classification = nc2.0,vi20   number of words = 2
partie parties

classification = ep0,lc0   number of words = 2
pendant voila'

classification = cs,dv7   number of words = 2
pourquoi quand

classification = an1.0   number of words = 2
premier second

classification = at1   number of words = 2
quel quels

classification = at2   number of words = 2
quelle quelles

classification = nc1.0,vi1,va1   number of words = 2
reste restes

classification = vi5,va5   number of words = 2
sont vont

classification = nc2.0,vn20   number of words = 2
sortie vue
```

```
classification = nc1.0,ep0   number of words = 2
sous vers

classification = aq2.0.1,pi8,ai   number of words = 2
toute toutes

classification = aq0.0.2,vi20   number of words = 2
venu venus

classification = vt2,va2   number of words = 1
a

classification = ep4,ep7   number of words = 1
a'

classification = nc1.0,vi0,va0   number of words = 1
aller

classification = vi3,va3   number of words = 1
allons

classification = ep0,dv1   number of words = 1
apre's

classification = vn21 pe8   number of words = 1
arre'tez-vous

classification = nc1.0,vt2,va2   number of words = 1
as

classification = vt21 pe7   number of words = 1
assieds-toi

classification = nc1.0,dv1   number of words = 1
aujourd

classification = dv0,cc   number of words = 1
aussi

classification = dv2,lc2   number of words = 1
autour

classification = ep0,dv0,lc0   number of words = 1
avant

classification = ep0,dv0   number of words = 1
avec

classification = nc1.0,vt7,va7   number of words = 1
avions
```

```
classification = aq1.0.0,nc1.0,dv2   number of words = 1
bas

classification = aq0.0.2,dv4,ko   number of words = 1
bien

classification = dv1,lc1   number of words = 1
biento't

classification = aq1.0.1,nc1.0,ko   number of words = 1
bon

classification = aq2.0.1,nc2.0   number of words = 1
bonne

classification = aq1.0.1,nc1.0   number of words = 1
bons

classification = pd13   number of words = 1
c

classification = pd12   number of words = 1
c*a

classification = nc1.0,cc   number of words = 1
car

classification = ad1,pe14   number of words = 1
ce

classification = pd1   number of words = 1
celui

classification = pd5   number of words = 1
celui-la'

classification = ad4   number of words = 1
ces

classification = ad2   number of words = 1
cet

classification = ad3   number of words = 1
cette

classification = pd9   number of words = 1
ceux-la'

classification = aq0.0.2,nc1.0,vt0   number of words = 1
cire'
```

```
classification = aq1.0.2,dv4   number of words = 1
clair

classification = aq1.0.0,nc1.0   number of words = 1
con

classification = nc0.0,vn1   number of words = 1
cours

classification = aq1.0.0,vn1   number of words = 1
court

classification = aq1.0.2,nc1.0,vt20   number of words = 1
couverts

classification = ep1,ep2,ep3,ep6   number of words = 1
d

classification = ep1,ep6   number of words = 1
de

classification = vt11   number of words = 1
de'moliront

classification = nc1.0,lc0,ep0   number of words = 1
de's

classification = aq1.0.0,lc1   number of words = 1
dernier

classification = ep3   number of words = 1
des

classification = vn21 pe15   number of words = 1
donne-le

classification = vi12   number of words = 1
dormant

classification = aq2.0.2,nc2.0,lc2   number of words = 1
droite

classification = ep2   number of words = 1
du

classification = aq0.0.2,vt20,va20   number of words = 1
du'

classification = ep0,vi1   number of words = 1
entre
```

```
classification = cs,ep0   number of words = 1
jusqu

classification = dn1,dn2,pe15,pe16   number of words = 1
l

classification = dn2,pe16   number of words = 1
la

classification = dv0,ko   number of words = 1
la'

classification = nc2.0,vt0   number of words = 1
lance'e

classification = pr6   number of words = 1
laquelle

classification = dn1,pe15   number of words = 1
le

classification = vn21 pe7   number of words = 1
le've-toi

classification = pr5   number of words = 1
lequel

classification = dn3,pe17   number of words = 1
les

classification = pe11   number of words = 1
lui

classification = ao4   number of words = 1
ma

classification = np2,nc2.0   number of words = 1
madame

classification = dv1,vt12   number of words = 1
maintenant

classification = cc,dv0   number of words = 1
mais

classification = aq1.0.0,nc1.0,dv4,lc1   number of words =
 1
mal
```

```
classification = pi8,ai,lc0   number of words = 1
me'me

classification = ao7   number of words = 1
mes

classification = vt21 pe17   number of words = 1
mets-les

classification = po6   number of words = 1
mienne

classification = aq1.0.0,dv4   number of words = 1
mieux

classification = aq2.0.2,vt20   number of words = 1
mise

classification = pe3   number of words = 1
moi

classification = nc2.0,lc2   number of words = 1
moitie'

classification = ao1   number of words = 1
mon

classification = aq1.0.2,nc0.0   number of words = 1
mort

classification = nc2.0,vm1   number of words = 1
neige

classification = aq1.0.2,nc1.0,ac   number of words = 1
neuf

classification = po4   number of words = 1
no'tre

classification = aq1.1.2   number of words = 1
normal

classification = pe4   number of words = 1
nous

classification = aq2.0.2,pi8,ai   number of words = 1
nulle

classification = aq0.0.0,vt0   number of words = 1
oblige's
```

```
classification = pi1   number of words = 1
on

classification = pr11,dv2   number of words = 1
ou'

classification = nc1.0,dv6   number of words = 1
pas

classification = aq0.0.2,nc1.0,vn0   number of words = 1
passe'

classification = vn1 pe5   number of words = 1
penses-tu

classification = nc2.0,pi5   number of words = 1
personne

classification = np1,nc1.0   number of words = 1
pierre

classification = vn21 pe17   number of words = 1
pose-les

classification = an2.0   number of words = 1
premie're

classification = pr4,dv0,cs   number of words = 1
qu

classification = pr2,dv0,cs   number of words = 1
que

classification = pi2   number of words = 1
quelqu

classification = pr1   number of words = 1
qui

classification = pr3   number of words = 1
quoi

classification = vt4 pe17   number of words = 1
remettez-les

classification = vt1,lc1   number of words = 1
renverse

classification = pi6   number of words = 1
rien
```

```
classification = pe13,cs   number of words = 1
s

classification = ao6   number of words = 1
sa

classification = pe13   number of words = 1
se

classification = vi11,va11   number of words = 1
seront

classification = ao9   number of words = 1
ses

classification = cs,dv5   number of words = 1
si

classification = nc1.0,ao3   number of words = 1
son

classification = nc2.0,vn5   number of words = 1
sorte

classification = nc2.0,lc0   number of words = 1
suite

classification = pe5,pe6   number of words = 1
t

classification = ao5   number of words = 1
ta

classification = pe6   number of words = 1
te

classification = ao8   number of words = 1
tes

classification = po2   number of words = 1
tien

classification = vn5,po[illegible]   number of words = 1
tienne

classification = vn1,po2   number of words = 1
tiens

classification = pe7   number of words = 1
toi
```

```
classification = ao2   number of words = 1
ton

classification = aq1.2.1,pi8,ai   number of words = 1
tous

classification = aq1.1.1,pi8,ai,dv0   number of words = 1
tout

classification = ep0,lc1   number of words = 1
travers

classification = an0.0   number of words = 1
troisie'me

classification = pe5   number of words = 1
tu

classification = in1   number of words = 1
un

classification = in2   number of words = 1
une

classification = vi21 pe6,va21 pe6   number of words = 1
va-t

classification = aq0.0.2,nc2.0,vi20   number of words = 1
venue

classification = aq1.0.2 ad   number of words = 1
vert-la'

classification = vt1 pe5,va1 pe5   number of words = 1
veux-tu

classification = aq0.0.2,vt1   number of words = 1
vide

classification = nc1.0,vn12   number of words = 1
volant

classification = vt4,va4   number of words = 1
voulez

classification = pe8   number of words = 1
vous

classification = vn3   number of words = 1
voyons
```

```
classification = pe20,dv2   number of words = 1
y
```

TABLE 39

Rank Summary of Lexical Categories

Frequency	Grammatical category
526	nc1.0
381	nc2.0
169	vt0
158	vt1
119	ms
89	vn0
85	vn1
75	ko
40	np2
40	nc1.0 ad
37	np1
37	aq0.0.2
35	aq0.0.2,vt0
31	nc2.0 ad
31	nc2.0,vt1
27	aq1.0.2
26	aq1.0.0
25	aq1.0.2,nc1.0
23	vi1
22	aq0.0.2,vn0
21	np0
17	aq2.0.0
15	dv0
15	vi0
15	aq2.0.2
12	nc1.0,vt1
12	ac,nc1.0
12	aq0.0.0
11	nc0.0
11	nc2.0,vn1
11	vn6
10	nc1.2
10	vt6
10	vt5

10	dv1
8	vt6,va6
8	vt10
8	vn9
8	ep0
7	vt1,va1
7	vi0,va0
7	nc1.0,lc0
7	vt9,va9
7	aq0.0.2,nc1.0
7	vt9
6	lc1
6	nc1.1
6	vn10
6	vt21 pe15
6	vt0,va0
6	nc2.0,vi1
6	aq0.0.2,nc2.0
6	vt20
6	vt5,va5
6	vi20
6	aq0.0.2,vt20
6	vi10
5	aq1.0.2,np1
5	aq0.0.2,vi0
5	vn20
5	vt21 pe3
5	vi1,va1
5	aq0.0.2,vn1
4	vi15,va15
4	dv2
4	dv3
4	vt15,va15
4	dv4
4	aq2.0.1
4	pi8,ai
4	nc1.0,dv2,ep0
4	vn5
4	vr1
4	vi2,va2
4	vt20,va20
4	lc0
4	vi10,va10
4	dv6
4	vi5
3	vn12
3	vt10,va10
3	aq0.0.1,pi8,ai
3	nc2.0,vn0
3	cs,dv0

3	nc1.0,vi0
3	vi6,va6
3	cc
3	vi9,va9
3	dv5
3	aq1.0.2,vn20
3	cs
3	aq2.0.2,nc2.0
2	nc1.0,vt12
2	nc2.0,vi0,va0
2	aq1.0.0,vt12
2	vi11
2	aq1.0.2,vt20
2	nc2.0,ko
2	ep5
2	vt11,va11
2	aq1.0.1
2	nc1.0,vi1
2	pd2
2	pd7
2	vt3
2	aq0.0.0,nc2.0
2	nc1.0,lc1
2	aq1.0.2,vt1
2	nc1.0,dv2,lc1
2	vn21 pe3
2	pe10
2	pe19,dv2,ep0
2	nc0.0 ad
2	vt12
2	lc3
2	vt21 pe3,va21 pe3
2	aq2.0.2,vt4,va4
2	vm1
2	pe9
2	pe1
2	aq2.0.2,np2
2	aq0.0.1
2	nc2.0,vt1,va1
2	pe18,po9,ao14
2	dv2,lc0
2	pe2
2	aq0.0.2,nc0.0
2	nc2.2
2	po1
2	vt21 pe16
2	aq2.0.2,vn20
2	nc2.0,vi20
2	ep0,lc0
2	cs,dv7

2	an1.0
2	at1
2	at2
2	nc1.0,vi1,va1
2	vi5,va5
2	nc2.0,vn20
2	nc1.0,ep0
2	aq2.0.1,pi8,ai
2	aq0.0.2,vi20
1	vt2,va2
1	ep4,ep7
1	nc1.0,vi0,va0
1	vi3,va3
1	ep0,dv1
1	vn21 pe8
1	nc1.0,vt2,va2
1	vt21 pe7
1	nc1.0,dv1
1	dv0,cc
1	dv2,lc2
1	ep0,dv0,lc0
1	ep0,dv0
1	nc1.0,vt7,va7
1	aq1.0.0,nc1.0,dv2
1	aq0.0.2,dv4,ko
1	dv1,lc1
1	aq1.0.1,nc1.0,ko
1	aq2.0.1,nc2.0
1	aq1.0.1,nc1.0
1	pd13
1	pd12
1	nc1.0,cc
1	ad1,pe14
1	pd1
1	pd5
1	ad4
1	ad2
1	ad3
1	pd9
1	aq0.0.2,nc1.0,vt0
1	aq1.0.2,dv4
1	aq1.0.0,nc1.0
1	nc0.0,vn1
1	aq1.0.0,vn1
1	aq1.0.2,nc1.0,vt20
1	ep1,ep2,ep3,ep6
1	ep1,ep6
1	vt11
1	nc1.0,lc0,ep0
1	aq1.0.0,lc1

```
1          ep3
1          vn21 pe15
1          vi12
1          aq2.0.2,nc2.0,lc2
1          ep2
1          aq0.0.2,vt20,va20
1          ep0,vi1
1          vr0
1          vi2 pe5,va2 pe5
1          pe12
1          vt21 pe15,va21 pe15
1          vt12,va12
1          aq1.0.2,vt1,va1,lc0
1          aq1.0.0,nc2.0
1          aq0.0.0,nc2.0,vn1
1          nc2.0,vt1,vm1
1          nc1.0,vn1
1          aq0.0.2,vn20
1          aq1.0.0,dv0
1          nc0.0,vt1
1          aq0.0.2,lc2
1          aq1.0.2,vt12
1          aq1.0.0,nc1.0,dv0,lc0
1          cs,ep0
1          dn1,dn2,pe15,pe16
1          dn2,pe16
1          dv0,ko
1          nc2.0,vt0
1          pr6
1          dn1,pe15
1          vn21 pe7
1          pr5
1          dn3,pe17
1          pe11
1          ao4
1          np2,nc2.0
1          dv1,vt12
1          cc,dv0
1          aq1.0.0,nc1.0,dv4,lc1
1          pi8,ai,lc0
1          ao7
1          vt21 pe17
1          po6
1          aq1.0.0,dv4
1          aq2.0.2,vt20
1          pe3
1          nc2.0,lc2
1          ao1
1          aq1.0.2,nc0.0
1          nc2.0,vm1
```

```
1          aq1.0.2,nc1.0,ac
1          po4
1          aq1.1.2
1          pe4
1          aq2.0.2,pi8,ai
1          aq0.0.0,vt0
1          pi1
1          pr11,dv2
1          nc1.0,dv6
1          aq0.0.2,nc1.0,vn0
1          vn1 pe5
1          nc2.0,pi5
1          np1,nc1.0
1          vn21 pe17
1          an2.0
1          pr4,dv0,cs
1          pr2,dv0,cs
1          pi2
1          pr1
1          pr3
1          vt4 pe17
1          vt1,lc1
1          pi6
1          pe13,cs
1          ao6
1          pe13
1          vi11,va11
1          ao9
1          cs,dv5
1          nc1.0,ao3
1          nc2.0,vn5
1          nc2.0,lc0
1          pe5,pe6
1          ao5
1          pe6
1          ao8
1          po2
1          vn5,po7
1          vn1,po2
1          pe7
1          ao2
1          aq1.2.1,pi8,ai
1          aq1.1.1,pi8,ai,dv0
1          ep0,lc1
1          an0.0
1          pe5
1          in1
1          in2
1          vi21 pe6,va21 pe6
1          aq0.0.2,nc2.0,vi20
```

1	aq1.0.2 ad
1	vt1 pe5,va1 pe5
1	aq0.0.2,vt1
1	nc1.0,vn12
1	vt4,va4
1	pe8
1	vn3
1	pe20,dv2

TABLE 40

Alphabetical Summary of Lexical Categories

Frequency	Grammatical category
12	ac,nc1.0
1	ad1,pe14
1	ad2
1	ad3
1	ad4
1	an0.0
2	an1.0
1	an2.0
1	ao1
1	ao2
1	ao4
1	ao5
1	ao6
1	ao7
1	ao8
1	ao9
37	aq0.0.2
12	aq0.0.0
2	aq0.0.1
35	aq0.0.2,vt0
22	aq0.0.2,vn0
5	aq0.0.2,vi0
5	aq0.0.2,vn1
1	aq0.0.2,lc2
1	aq0.0.0,vt0
1	aq0.0.2,vt1

6	aq0.0.2,vt20
2	aq0.0.2,vi20
1	aq0.0.2,vn20
7	aq0.0.2,nc1.0
6	aq0.0.2,nc2.0
2	aq0.0.0,nc2.0
2	aq0.0.2,nc0.0
3	aq0.0.1,pi8,ai
1	aq0.0.2,dv4,ko
1	aq0.0.2,nc1.0,vt0
1	aq0.0.2,vt20,va20
1	aq0.0.0,nc2.0,vn1
1	aq0.0.2,nc1.0,vn0
1	aq0.0.2,nc2.0,vi20
27	aq1.0.2
26	aq1.0.0
2	aq1.0.1
1	aq1.0.2 ad
5	aq1.0.2,np1
2	aq1.0.2,vt1
1	aq1.0.2,dv4
1	aq1.0.0,vn1
1	aq1.0.0,lc1
1	aq1.0.0,dv0
1	aq1.0.0,dv4
3	aq1.0.2,vn20
2	aq1.0.0,vt12
2	aq1.0.2,vt20
1	aq1.0.2,vt12
25	aq1.0.2,nc1.0
1	aq1.0.1,nc1.0
1	aq1.0.0,nc1.0
1	aq1.0.0,nc2.0
1	aq1.0.2,nc0.0
1	aq1.0.1,nc1.0,ko
1	aq1.0.2,nc1.0,ac
1	aq1.0.0,nc1.0,dv2
1	aq1.0.2,nc1.0,vt20
1	aq1.0.2,vt1,va1,lc0
1	aq1.0.0,nc1.0,dv0,lc0
1	aq1.0.0,nc1.0,dv4,lc1
1	aq1.1.2
1	aq1.1.1,pi8,ai,dv0
1	aq1.2.1,pi8,ai
17	aq2.0.0
15	aq2.0.2
4	aq2.0.1
2	aq2.0.2,np2
2	aq2.0.2,vn20
1	aq2.0.2,vt20

3	aq2.0.2,nc2.0
1	aq2.0.1,nc2.0
2	aq2.0.1,pi8,ai
1	aq2.0.2,pi8,ai
2	aq2.0.2,vt4,va4
1	aq2.0.2,nc2.0,lc2
2	at1
2	at2
3	cc
1	cc,dv0
3	cs
3	cs,dv0
2	cs,dv7
1	cs,dv5
1	cs,ep0
1	dn1,dn2,pe15,pe16
1	dn1,pe15
1	dn2,pe16
1	dn3,pe17
15	dv0
1	dv0,cc
1	dv0,ko
10	dv1
1	dv1,lc1
1	dv1,vt12
4	dv2
2	dv2,lc0
1	dv2,lc2
4	dv3
4	dv4
3	dv5
4	dv6
8	ep0
1	ep0,dv1
1	ep0,dv0
1	ep0,dv0,lc0
2	ep0,lc0
1	ep0,lc1
1	ep0,vi1
1	ep1,ep6
1	ep1,ep2,ep3,ep6
1	ep2
1	ep3
1	ep4,ep7
2	ep5
1	in1
1	in2
75	ko
4	lc0
6	lc1

2	lc3
119	ms
11	nc0.0
2	nc0.0 ad
1	nc0.0,vn1
1	nc0.0,vt1
526	nc1.0
40	nc1.0 ad
1	nc1.0,cc
12	nc1.0,vt1
7	nc1.0,lc0
3	nc1.0,vi0
2	nc1.0,vi1
2	nc1.0,lc1
2	nc1.0,ep0
1	nc1.0,dv1
1	nc1.0,vn1
1	nc1.0,dv6
1	nc1.0,ao3
2	nc1.0,vt12
1	nc1.0,vn12
4	nc1.0,dv2,ep0
2	nc1.0,dv2,lc1
2	nc1.0,vi1,va1
1	nc1.0,vi0,va0
1	nc1.0,vt2,va2
1	nc1.0,vt7,va7
1	nc1.0,lc0,ep0
6	nc1.1
10	nc1.2
381	nc2.0
31	nc2.0 ad
2	nc2.0,ko
31	nc2.0,vt1
11	nc2.0,vn1
6	nc2.0,vi1
3	nc2.0,vn0
1	nc2.0,vt0
1	nc2.0,lc2
1	nc2.0,vm1
1	nc2.0,pi5
1	nc2.0,vn5
1	nc2.0,lc0
2	nc2.0,vi20
2	nc2.0,vn20
2	nc2.0,vi0,va0
2	nc2.0,vt1,va1
1	nc2.0,vt1,vm1
2	nc2.2
21	np0

37	np1
1	np1,nc1.0
40	np2
1	np2,nc2.0
1	pd1
1	pd12
1	pd13
2	pd2
1	pd5
2	pd7
1	pd9
2	pe1
2	pe10
1	pe11
1	pe12
1	pe13
1	pe13,cs
2	pe18,po9,ao14
2	pe19,dv2,ep0
2	pe2
1	pe20,dv2
1	pe3
1	pe4
1	pe5
1	pe5,pe6
1	pe6
1	pe7
1	pe8
2	pe9
1	pi1
1	pi2
1	pi6
4	pi8,ai
1	pi8,ai,lc0
2	po1
1	po2
1	po4
1	po6
1	pr1
1	pr11,dv2
1	pr2,dv0,cs
1	pr3
1	pr4,dv0,cs
1	pr5
1	pr6
15	vi0
7	vi0,va0
23	vi1
5	vi1,va1
6	vi10

```
  4          vi10,va10
  2          vi11
  1          vi11,va11
  1          vi12
  4          vi15,va15
  1          vi2 pe5,va2 pe5
  4          vi2,va2
  6          vi20
  1          vi21 pe6,va21 pe6
  1          vi3,va3
  4          vi5
  2          vi5,va5
  3          vi6,va6
  3          vi9,va9
  2          vm1
 89          vn0
 85          vn1
  1          vn1 pe5
  1          vn1,po2
  6          vn10
  3          vn12
  5          vn20
  2          vn21 pe3
  1          vn21 pe8
  1          vn21 pe7
  1          vn21 pe15
  1          vn21 pe17
  1          vn3
  4          vn5
  1          vn5,po7
 11          vn6
  8          vn9
  1          vr0
  4          vr1
169          vt0
  6          vt0,va0
158          vt1
  1          vt1 pe5,va1 pe5
  1          vt1,lc1
  7          vt1,va1
  8          vt10
  3          vt10,va10
  1          vt11
  2          vt11,va11
  2          vt12
  1          vt12,va12
  4          vt15,va15
  1          vt2,va2
  6          vt20
  4          vt20,va20
```

5	vt21 pe3
1	vt21 pe7
6	vt21 pe15
2	vt21 pe16
1	vt21 pe17
2	vt21 pe3,va21 pe3
1	vt21 pe15,va21 pe15
2	vt3
1	vt4 pe17
1	vt4,va4
10	vt5
6	vt5,va5
10	vt6
8	vt6,va6
7	vt9
7	vt9,va9

REFERENCES

Dubois, J. *Grammaire structurale du français: nom et pronom*. Paris: Larousse, 1965.

Dubois, J., & Dubois-Charlier, F. *Elements de linguistique française: syntaxe*. Paris: Larousse, 1970.

Gammon, E. M. A syntactical analysis of some first-grade readers. Technical Report No. 155, June 22, 1970, Stanford University, Institute for Mathematical Studies in the Social Sciences. Reprinted in K. J. Hintikka, J. Moravcsik & P. Suppes (Eds.), *Approaches to natural language*. Dordrecht, Holland: Reidel, in press.

Grévisse, M. *Le bon usage, Grammaire française avec des remarques sur la langue française d'aujourd'hui*. Gemblouz, France: Duculot, 1969.

Martinet, M. A. De l'économie des formes du verbe en français parlé. A. G. Hatcher & K. L. Selid (Eds.), *Studia philologica et litteraria in honorem L. Spitzer*. Bern, Switzerland: Francke, 1958.

Robert, P. *Le petit Robert*. Paris: Société du nouveau Littré, 1967.

Smith, R. L., Jr. The syntax and semantics of ERICA. Technical Report No. 185, June 14, 1972, Stanford University, Institute for Mathematical Studies in the Social Sciences.

Suppes, P. Probabilistic grammars for natural languages. Technical Report No. 154, May 15, 1970, Stanford University, Institute for Mathematical Studies in the Social Sciences. Reprinted in *Synthese*, 1970, *22*, 95-116.

Suppes, P. Semantics of context-free fragments of natural languages. Technical Report No. 171, March 30, 1971, Stanford University, Institute for Mathematical Studies in the Social Sciences. Reprinted in K. J. Hintikka, J. Moravcsik & P. Suppes (Eds.), *Approaches to natural language*. Dordrecht, Holland: Reidel, in press.

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
Institute for Mathematical Studies in the
Social Sciences, Stanford University

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
The French Syntax and Semantics of Philippe, Part 1: Noun Phrases

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(First name, middle initial, last name)*
P. Suppes, R. Smith, and M. Léveillé

6. REPORT DATE
11-3-72

7a. TOTAL NO. OF PAGES
191

7b. NO. OF REFS
9

8a. CONTRACT OR GRANT NO.
N0014-67-0112-0049

b. PROJECT NO.
NR 049-303

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
Technical Report No. 195

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT
Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Off. of Naval Res., Dept. of the Navy
Information Systems Program, Code 437
Arlington, Va. 22217

13. ABSTRACT

The details of the collection of a 33-hour corpus of French psycholinguistic data are given, including a methodological analysis of the purposes of the investigation. Simple measures, such as the length of utterance, and the size and kind of vocabulary, are applied to the corpus by way of introduction.

As a prelude to further work on the syntax and semantics of this corpus, two grammars for the noun phrases within a fragment of the corpus are discussed. Particular emphasis is placed upon the relation of this work to previous works regarding English corpora. Probabilistic fits of these grammars are obtained from these grammars via computer programs, and the developmental aspects of these fits are discussed.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6801

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
psycholinguistics						
syntax						
set-theoretical semantics						
probabilistic grammars						

Security Classification